

Result	Query	
No.	Score	Match Length DB ID Description
-----		

1	240.6	27.9	1022	8	ABO05888	Arabidops	
2	237	27.5	1033	8	PSMYB26	Y1105 P. sativum m	
3	231.8	26.9	645	8	AF175987	AF175987 Arabidops	
4	223	25.9	872	8	ABO58642	ABO58642 Liliuim hy	
5	192.8	22.4	1343	8	AY026332	AY026332 Oryza sat	
6	176.4	20.5	1295	8	CPU3173	U33917 Craterostig	
7	167.8	19.5	1423	8	AF262733	AF262733 Arabidops	
8	166.4	19.3	133242	8	AP003140	AP003140 Oryza sat	
9	166.4	19.3	145491	8	AP002883	AP002883 Oryza sat	
10	162.2	18.8	1140	8	AF334815	AF334815 Arabidops	
11	155.4	18.0	931	8	AT006377	AT006377 Arabidops	
12	151.4	17.6	1033	8	AF034133	AF034133 Gossypium	
13	143.2	16.6	1513	8	AF427146	AF427146 Zea mays	
14	143.2	16.6	1601	8	ZMU57002	ZMU57002 Zea mays P	
15	143.2	16.6	1176	8	MEEPR	M73028 Zea mays pr	
16	138.2	15.0	1176	8	D88617	D88617 Oryza sativ	
17	136.8	15.9	1202	8	OSMYB1202	Y1114 O. sativa mR	
18	136.8	15.9	1270	8	HMYB3	X70878 H. vulgare m	
19	136.2	15.8	2162	8	PPPSRNA	X67050 P. patens mR	
20	134.8	15.6	489	8	AB044084	AB044084 Trillium	
21	134.2	15.6	1024	8	AF062894	AF062894 Arabidops	
22	133.8	15.5	1070	8	ATU62743	U62743 Arabidops	
23	133.8	15.5	125599	2	AF003815	AF003815 Oryza sat	
24	133.8	15.5	13531	2	AF003813	AF003813 Oryza sat	
C	25	133.6	15.5	2352	6	AR164405	AR164405 Sequence
	26	133.6	15.5	2373	8	OSMYB	X98355 O. sativa mR
	27	133.4	15.5	1032	8	AF371974	AF371974 Arabidops
	28	133.4	15.5	1069	8	D88618	D88618 Oryza sativ
	29	133.4	15.5	1512	8	AY063939	AY063939 Arabidops
	30	132.8	15.4	1050	8	AF175989	AF175989 Arabidops
	31	132.8	15.4	1235	8	AT060588	AT060588 Arabidops
	32	132.8	15.4	1284	8	HMYB1	X70877 H. vulgare m
	33	132.8	15.4	1334	8	AF062864	AF062864 Arabidops
	34	132.6	15.4	1102	8	HMYB2	X70876 H. vulgare m
35	131.6	15.3	2287	8	ASAI33658	ASAI33658 Avena sat	
36	131	15.2	1220	8	AF161711	AF161711 Pimpinell	
37	130.6	15.2	819	8	ABO29165	ABO29165 Glycine m	
38	130.4	15.1	1869	8	AF114162	AF114162 Lolium te	
39	130.2	15.1	1259	8	AF249310	AF249310 Arabidops	
40	130.2	15.1	1355	8	OSMYB1355	Y11495 O. sativa mR	
41	130	15.1	2220	6	AR164404	AR164404 Sequence	
42	130	15.1	2263	8	HVRNAGAM1	X87690 H. vulgare m	
43	128.6	14.9	795	8	AB029160	AB029160 Glycine m	
44	128.6	14.9	1135	8	AF401220	AF401220 Fragaria	
45	128	14.8	1606	8	AF034132	AF034132 Gossypium	

## ALIGNMENTS

RESULT	1
LOCUS	AB005888
DEFINITION	Arabidopsis thaliana mRNA for ATMYB3, complete cds.
ACCESSION	AB005888
VERSION	AB005888.1 GI:2280527
KEYWORDS	ATMYB3.
SOURCE	Arabidopsis thaliana (strain:Columbia) cDNA to mRNA.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1033)
REFERENCE	Noji,M.
AUTHORS	Direct Submission
TITLE	Submitted (18-JUL-1997) Masaaki Noji, Chiba University, Faculty of
JOURNAL	Pharmaceutical Sciences; Yayoi-cho 1-33, Inage-Ku, Chiba, Chiba
	263, Japan (E-mail:mmoji@p.chiba-u.ac.jp, Tel:+81-43-290-2906, Fax:+81-43-290-2905)
REFERENCE	2 (sites)
AUTHORS	Noji,M., Urao, T., Shinozaki,K.Y. and Shinozaki,K.
TITLE	Molecular cloning of two cDNAs encoding novel myb homologs from
	Arabidopsis (Accession Nos. AB005888 and AB005889) (Pc998-111)

JOURNAL FEATURES	Plant Physiol.	117, 720 (1998)
source	Location/Qualifiers	
gene	1. 1033	
CDS	/organism="Arabidopsis thaliana"	
	/strain="Columbia"	
	/db.xref="taxon:3702"	
	115. .795	
	/gene="Atmyb3"	
	115. .795	
	/gene="Atmyb3"	
	/codon_start=1	
	/product="ATMYB3"	
	/protein_id="BAA21618.1"	
	/db.xref="GI:2280528"	
	/translation="MEKRGCGSSCGGSSAFAEYRKGPWTWEEDLLILINYIANHGDEV WNSLSAGLKRKTSGLRLMLNLRPDVRGNTTPEQLIMELHAKGNRWSIKRK HLPTDNEIKNFCRTLRQKIRKSDVYTTSSVGHSSSEINDQASTSSSHNVCTDD QAMETSYPTPSYGHRTNMEFNYSNAATATATVDYVPKTIYVDQGTENYWGMDINS SMHLNCG"	
BASE COUNT	362 a	208 g
ORIGIN	182 c	281 t

RESULT	2
LOCUS	PSMYR26
DEFINITION	P.sativum mRNA for Myb-like protein (Myb26).
ACCESSION	Y11105
VERSION	Y11105.1 GI:1841474
KEYWORDS	Myb-like protein; Myb26.
SOURCE	pea.
ORGANISM	Plum sativum
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Plum.
REFERENCE	1 (bases 1 to 1022)
AUTHORS	Ulmairi, A. and Strommer, J.
TITLE	Myb26: a MYB-like protein of pea flowers with affinity for promoters of phenylpropanoid genes
JOURNAL	Plant J. 12 (6), 1273-1284 (1997)
MEDLINE	98112025
REFERENCE	2 (bases 1 to 1022)
AUTHORS	Strommer, J.N.

TITLE	Journal	Submitted	Direct Submission
FEATURES	Molecular Biology And Genetics, Guelph, Ontario N1G 2W1, CANADA	05-FEB-1997	J.N. Strommer, University Of Guelph,
SOURCE	Location/Qualifiers	1. .1022	
	/organism="Pisum sativum"		
	/db_xref="taxon:3888"		
	/tissue_type="flower bud"		
CDS	92. .745		
	/codon_start=1		
	/product="Myb26"		
	/protein_id="CA71992.1"		
	/db_xref="GI:1841475"		
	/db_xref="SPTRMBL:P93474"		
	/translation="MDKKPCNSSDPPEVRKGPWTMEDLLILINYIANHGCVNWSLAK AAGLKRTGSCRLRWLYLRDVRGRNITFEEDLLIMELHSKMGNRSSAKHLPGPT DNEIKNFWRTLOKHRYOVNDPQNOQFOOKSLLEINDHNHNHNPSSOVSNLYEPM ETYSPTSGCLTEPPEPPTIINDHNHNSCCANDNNNNNYSMEDIMWMLLNGD"		
BASE COUNT	370 a 170 c 161 g 321 t		
ORIGIN			

AF175987	645 bp	mRNA	linear	PLN 30-AUG-2001
LOCUS	Arb1dops1s thallana putative transcription factor (MYB24) mRNA,			
DEFINITION	complete cds.			
ACCESSION	AF175987			
VERSION	AF175987.1	GI:5823306		
KEYWORDS				
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsids.			
REFERENCE	1 (bases 1 to 645)			
AUTHORS	Kiranz, H.D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meisner, R.C.,			
	Petroni, K., Urzainqui, A., Beyer, M., Martin, C., Smeekens, S.,			
	Tonelli, C., Paz-Ares, J. and Weishaar, B.			
TITLE	Towards functional characterisation of the members of the R2R3-MYB			
JOURNAL	gene family from Arabidopsis thaliana			
	The plant journal : for cell and molecular biology. 16 (2), 263-276			
	(1998)			
MEDLINE	99056848			

PUBMED 9839469  
 REFERENCE 2 (bases 1 to 645)  
 AUTHORS Stracke,R., Werber,M. and Weisshaar,B.  
 TITLE The R2R3-MYB gene family in Arabidopsis thaliana  
 JOURNAL Curr. Opin. Plant Biol. 4, 447-456 (2001)  
 REFERENCE 3 (bases 1 to 645)  
 AUTHORS Stracke,R. and Weisshaar,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-1999) Dept. Biochemie, Max-Planck-Institut fuer Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln D-50829, Germany

FEATURES  
 source  
 1..645  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /note="ecotype: Columbia"  
 1..645  
 /gene="MYB24"  
 1..645  
 /gene="MYB24"  
 /note="R2R3-MYB family member"  
 /codon\_start=1  
 /product="putative transcription factor"  
 /protein\_id="AAD53092.1"  
 /db\_xref="GI:5823307"  
 /translation="MEKRESGGSGDAEVRKGPMTMEDLLINTYANHGEGVANS  
 LAKSAGIKRTGSKRLNLYRPVRRGNITPREQLIMELHQRGWRWSKIAKHL  
 GRTDNEIKRFRKTKIOKTIKSGETTTGSSOSEPTNIHHTSHVANDTOETMDYSP  
 TTSYOHASHINQDLNNGNYVPSGSTIMPELVSQSEONWVSVDLPMNITNGN"

BASE COUNT 219 a 125 c 153 g 148 t  
 ORIGIN

Query Match 26.9%; Score 231.8; DB 8; Length 645;  
 Best Local Similarity 81.3%; Pred. No. 9.7e-55;

Matches 269; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 64 tcccaagatcgaagtgagaaagggccgtgagcatgtaagaagatcctccatc 123  
 DB 34 TCAGGAGATCGAAGTGAAGAAAGGCCATGACATGAGAAAGATTGATTCATC 93  
 QY 124 aactacaatagaacacgagcggaagcagttggaactccctgacaaagctgctgctca 183  
 DB 94 AATTATATCGCAATCATGTTGGAAGGTTGGAACCTCTCGCAAAATCTGCAGAGCTA 153  
 QY 184 aaacgtacccgggaaggttgcgtccggtggtgtaactatctgaccccgccg 243  
 DB 154 AAACGACCGGGAAGAGTTCCCGCTCCGGTGGCTGAACCTCCGACCTGATGCGCA 213  
 QY 244 agaggaacaactcaactatgagagcagctcctgatactgtaactgcatgccaagtggga 303  
 DB 214 CGGGGAATATTCACACCAAGAACACATCATGTAATCTGCAAAATGAGGA 273  
 QY 304 aacaggtggtcctaaatltgcaaacgactcctccggaagacgtgacaatgataaagac 363  
 DB 274 AATAGGTGTCACAAATATTCGCAAAAGCATTTACAGGAAGACGACATGAGATAAGAAAT 333  
 QY 364 ttctggagagatgataatcccaaaagacatca 394  
 DB 334 TTTTGGAGACTAAGATCCAGAAATACATCA 364

RESULT 4  
 AB058642 872 bp mRNA linear PLN 04-APR-2001  
 LOCUS AB058642  
 DEFINITION Lilium hybrid division I Lmhyb mRNA, complete cds.  
 ACCESSION AB058642  
 VERSION AB058642.1 GI:13537529  
 KEYWORDS  
 SOURCE Lilium hybrid division I (cultivar:Montreux) tepal cDNA to mRNA.  
 ORGANISM Lilium hybrid division I  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae;  
 Lilium; Lilium hybrid cultivars.

REFERENCE 1 (sites)  
 AUTHORS Nakatsuka,A., Izumi,Y. and Yamagishi,M.  
 TITLE Isolation and characterization of the genes related to anthocyanin biosynthesis in Asiatic hybrid lily  
 JOURNAL unpublished  
 REFERENCE 2 (bases 1 to 872)  
 AUTHORS Nakatsuka,A., Izumi,Y. and Yamagishi,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAR-2001) Akira Nakatsuka, Shimane University, Faculty of Life and Environmental Sciences, Nishiwatsu 1060, Matsue, Shimane 690-8504, Japan  
 (E-mail:nakira@life.shimane-u.ac.jp, Tel:01-852-32-6502)

FEATURES  
 source  
 1..872  
 /organism="Lilium hybrid division I"  
 /cultivar="Montreux"  
 /db\_xref="taxon:156532"  
 /tissue\_type="tepals"  
 80..634  
 /gene="Lmhyb"  
 80..634  
 /gene="Lmhyb"  
 /codon\_start=1  
 /protein\_id="BAB40790.1"  
 /db\_xref="GI:13537530"  
 /translation="MDRKVGTNEDNVEVRKGPMTMEDLLINTYANHGEGVANS  
 AGLRKRTGSKRLNLYRPVRRGNITPREQLIMELHQRGWRWSKIAKHL  
 NEIKNFWTRVOKRSKNSESESGOISMIDETSAMOHEDVDVGSOTSDOGOTSNML  
 DAFELPEEPGNLLSVELWMOQ"

BASE COUNT 269 a 177 c 213 g 213 t  
 ORIGIN

Query Match 25.9%; Score 223; DB 8; Length 872;  
 Best Local Similarity 69.2%; Pred. No. 3.2e-52;

Matches 304; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 21 caagctgacatcggagaaagccagagcagcagtgtaagtcacaaagtgcaggt 80  
 DB 58 CATGATGAGGAGAGAACACACCATGAGACAGAAAGTAGGACTAATGAAGATTGGAGCT 117  
 QY 81 gagaagaagccgttgagcagtggaagagatctcatcctcaactacatagcgaatca 140  
 DB 118 GAGGAAGGACCATGAGACCATGAGAGAGATCTCATCTCATCAACTACATAGCCACCA 177  
 QY 141 cggcgaagcagttggaactccctgacaaagctgctggtcctaaagctacccggaagag 200  
 DB 178 TGGCGACGGAGTCTGGAMACACGCTGCTGATCGGCGGACTGAAGAGACTGGAAAGAG 237  
 QY 201 ttgtcgctccggttgctgaactatctgacccagcagctcggagaggaacatactac 260  
 DB 238 CTGCAGGCTTCTGTGGCTTAATTACCTGAGGCTGATGTTCCTCGGGGTACATACACCC 297  
 QY 261 tgaagacagctcctcgtatcatctggaactgacatgccaagtggggaacaggtggtcctaaat 320  
 DB 298 AGAGAGACAGCTTCTGATCATGGAATCTCACAGATGCGGTGACAGATGCTGCAAAAT 357  
 QY 321 tgcgaagcactcctccggaagagctgacaatgagataaagaactcttggaagcagaat 380  
 DB 358 TGCAGGAGGCTCCACGAGACGAAAGATTAAGCAAGAACTCTTGAGAGACCGAGAT 417  
 QY 381 ccaaaagcatalcaagcaagcagagccttctcgttcagagctccggaatgagtagatca 440  
 DB 418 CCGAAGGAATTCACAAATGCGCAATCATCTGAGGGCCAAATCTCATCTTTCGATGA 477  
 QY 441 agcaagcacaaagccacatg 459  
 DB 478 AACTAGTCCCATGCAGATG 496

RESULT 5  
 AY026332 1343 bp mRNA linear PLN 30-JAN-2002  
 LOCUS AY026332

DEFINITION	ORYZA SATIVA MYB TRANSSCRIPTION FACTOR JAMYB mRNA, complete cds.
ACCESSION	AY026332
VERSION	AY026332.1
KEYWORDS	GI:13177577
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa. Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1 (bases 1 to 1343)
AUTHORS	Lee, M.-W., Qi, M. and Yang, Y.
TITLE	A novel jasmonic acid-inducible rice myb gene associates with fungal infection and host cell death
JOURNAL	Mol. Plant Microbe Interact. 14 (4), 527-535 (2001)
MEDLINE	11310458
PUBMED	11310740
REFERENCE	2 (bases 1 to 1343)
AUTHORS	Lee, M.-W. and Yang, Y.
TITLE	Direct Submission
JOURNAL	Submitted (29-JAN-2001) Plant Pathology, University of Arkansas, 217 Plant Science Bldg., Fayetteville, AR 72701, USA
FEATURES	Location/Qualifiers
source	1..1343
CDS	/organism="Oryza sativa" /cultivar="drew" /db_xref="taxon:4530" 127..984 /note="related to host cell death and defense responses; induced by jasmonic acid, wounding, or infection of rice blast fungus, but not by salicylic acid or abscisic acid" /codon_start=1 /product="Myb transcription factor JAMYB" /protein_id="AAK08983.1" /db_xref="GI:13177578" /translation="MEMVLTQSHHPVPGEDQEAALFSAELRGPTVDELTITN YISGSGRMALAKAGIKRTGSCSLRYNTXRPVYKGNPFAEDLLDLDSRK GNRSKTAQHLPGFTDNEIKRNTYTRQKNAKQACDYSNRRFDANKLMPRLAEK IHAKGAVDDSGDISNDLSKVCSTYATVATACDGSFMSYSSDSFSEODLRK INLVHMDDEKMSDEMDQEVDFEFWSTETIOPNNPQDQDLGNWVGFSGLSETLLM SLDEIMQD"
BASE COUNT	367 a 298 c 337 g 341 t
ORIGIN	
Query Match	22.4%: Score 192.8; DB 8; Length 1343;
Best Local Similarity	74.7%: Pred. No. 1.3e-43;
Matches	242; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY	76 gaagtgaagaaagggcggtgagcagatgaagaagagatcattccatcaactaactaagc 135
DB	208 GAGCTCCGGCGAGGGCCCTGTGACCTCGACGAGGACCTCACCCTCATCAATTAATCTCT 267
QY	136 aatcacgagaaagcagctgtgaactcccttaagccaaagctgtcgtctaaacgtaccgg 195
DB	268 GATCAGCGCGAGGGCCGTGTGAACCACTTCGACACCGCCGCCGCTGTGAAGAGGACTGG 327
QY	196 aagaattgtggctccgggtggtcgaactacatctcgagcccgacgctccggaagagaaact 255
DB	328 AAGACCTCCGGCTCCGTGCTGAACATATCTCGCGGATGTAAAGCGCGCAACTTC 387
QY	256 actactgaagacagctccatgatactgaactgaatgcacaaagtggggaacacagttgct 315
DB	388 ACCCGAGGAGGACACTCTCTCATCTCCACACCTCCACTCCCGATGGGCAACGAGTGCC 447
QY	316 aaattgaagaagactcccggaagagactgaatgaatgaagaagaactctcgagagact 375
DB	448 AAGATGACCAACAATTTCCTGTGGAGGACGACCAACGACGATCAACAACTACTGAGGAGC 507
QY	376 agaattcaaaagacataagca 399
DB	508 AGAGTGCAAAAAGCATGCCCAACAA 531

CPUS3917	LOCUS DEFINITION	Accession version	KEYWORDS	SOURCE ORGANISM	TITLE	JOURNAL MEDLINE REFERENCE AUTHORS	FEATURES
CPUS3917	CPU33917	U33917.1	GI:1002799	Craterostigma plantaginellum.	A family of novel myb-related genes from the resurrection plant	Journal of Molecular Biology	Location/Occurrence
LOCUS DEFINITION	Craterostigma plantaginellum myb-related transcription factor (cpm7)			Craterostigma plantaginellum.	Craterostigma plantaginellum.	Craterostigma plantaginellum.	Craterostigma plantaginellum.
Accession version	U33917	U33917.1	GI:1002799	Eukaryota; Viridiplantae;	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Lamiales; Lamiales Incertae sedis; Totoniaceae; Craterostigma.	1 (bases 1 to 1295)	1 (bases 1 to 1295)
Source organism	Craterostigma plantaginellum.			Craterostigma plantaginellum.	Craterostigma plantaginellum.	Craterostigma plantaginellum.	Craterostigma plantaginellum.
Title	A family of novel myb-related genes from the resurrection plant			Craterostigma plantaginellum are specifically expressed in callus and roots in response to ABA or desiccation	Plant Mol. Biol. 32 (4), 707-716 (1996)	97134962	2 (bases 1 to 1295)
Journal medline reference authors	Iturriga, G., Leyns, L., Villegas, A., Gharaibeh, R., Salamini, F. and Batteals, D.						
Features	Direct Submission			Instituto de Biotecnologia, Av. Universidad #2001, Cuernavaca, MOR 62210, Mexico			
Source	1. 1295			/organism="Craterostigma plantaginellum"			
gene	1. .1295			/db_xref="taxon:4153"			
CDS	125..1132			/gene="cpm7"			
	125..1132			/gene="cpm7"			
	/note="putative DNA-binding protein; Description: myb-like gene: myb-related transcription factor"			/codon_start=1			
	/product="Cpm7"			/protein_id="AA858314.1"			
	/db_xref="GI:1002800"			/translation="MNOQVVVKKSKNQINSEDDSSDLRGFWYDEDFTLINITYA			
	HGGGRNNSLRFAFGALRKYKTSRCRLMYLIRPVRRKNTITLEDQLILSHSGNR			WSKAIOHLCPTDMEIKRWRTROKHAQOLKCVNSKOFEDTRYLMPLVRIGOR			
	AATVDGAPPLASSASSPSSAMNTACYSTAMAADHRQFMLMPQYAMTHSNLMTIA			QENSIVASSSEFCGLSELSELTANYANHVGINADHQOISSTTSGMONDEPAINT			
	VGNNGNSDQGLMGSVDDRRSNEGMWMDVDVNGSSPDKNLLMNVDVMTLQGFSSC			F"			
BASE COUNT	405 a	286 c	279 g	325 t			
ORIGIN							
Query Match	20.5%	Score 176.4	DB 8	Length 1295			
Best Local Similarity	70.9%	Pred. No. 5.9e-39					
Matches 234	Conservative 9%	Mismatches 96%	Indels 0%	Gaps 0%			
Oy	76 gaggtgagaaaaaggccgtgacagcttggaagaagtcatacctcatgaactatagcg	135					
Dd	197 GACCTCAAGAAGAGCGCCGTGACTTTATGAAGACTTCACACTCATACATCGCT	256					
Oy	136 aatcacggcgaagcagtgtgaacctccttagccaagctgcgtgctaaacgtaccggg	195					
Dd	257 CACCATTGCCGAAGGAATGGAACCTCTTGACGCTTTCGCCTGGAAGAACACTCGA	316					
Oy	196 aagaatttgcgtcgcgtgtgctgaacatcttgcaaccgaagctcggagaggaacact	255					
Dd	317 AAGACCTGCACATTGATGATGCTTAACACTACTTGAGACCCGAGTCTTCGTCGAGGAATATTC	376					
Oy	256 actactgagagcagctccctgatcatgtgaactgcacgcaatvgggagaacagttgctc	315					
Dd	377 ACACCTGGAAGAACAACCTTCTCATTTCTTAACCTCCATTTCTGTTGGGGCAATAAGGTGGC	436					

[illegible]

RESULT	7
AF262733	
LOCUS	1423 bp mRNA linear PLN 30-AUG-2001
DEFINITION	Arabidopsis thaliana putative transcription factor MYB108
ACCESSION	AF262733 complete cds.
VERSION	AF262733.2 GI:15375290

**SOURCE ORGANISM**

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
1 (bases 1 to 1423)  
AUTHORS  
Stracke, R., Werber, M. and Weisshaar, B.  
TITLE  
The R3R3-MWB gene family in *Arabidopsis thaliana*  
JOURNAL  
Curr. Opin. Plant Biol. 4 (5), 447-456 (2001)  
ISSN  
1469-3700

MEDLINE	21481677
PUBMED	11597504
REFERENCE	2 (bases 1 to 1423)
AUTHORS	Stracke, R. and Weisshaar, B.

TITLE	Direct submission
JOURNAL	Submitted (02-MAY-2000) Biochemie, Max-Planck-Institut fuer Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln D-50829, Germany
REFERENCE	3 (bases 1 to 1423)

**AUTHORS** Stracke, R. and Weissnaar, B.  
**TITLE** Direct Submission  
**Submitted (29-AUG-2001)** Dept. Plant Breeding and Yield Physiology,  
Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg

REMARK	10, Koeln D-50829, Germany
COMMENT	Sequence update by submitter
FEATURES	On Aug 30, 2001 this sequence version replaced gi:8101955.
	Location/Qualifiers

```
source
1..1423
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
```

```

/note="ecotype: Columbia
gene
1. .1423
/gene="MYB108"
1. .154
5'UTR

```

```
misc_feature    /gene="MYB108"  
                5  
                /gene="MYB108"  
                /note="alternative transcription start site"
```

CDS 155..1126  
/gene="MYB108"  
/note="At3g06490; member of the R2R3-MYB family  
/codon\_start=1

```

/product="putative transcription factor MYB108"
/protein_id="AA072668.1"
/db_xref="GI:8101956"
/translation="MDKGRSLKNNMEDMDLRGPTAAEDFKLNNYIATNGSGRR

```

NSLSRQAGLORTGKSCRLRWLNITLPDVRKRNITLEEDLITLHLSRQGNMKSITAOQ  
LPSTDIIEIKNYMRTROKHAKOLKCDVNSOFPOTMQLVLMNPRLVEITIOSASASA  
AATTTTTTTSAGTSSCITTSNNOFNMTDINNMMGOQFEVMSNDYITPENNVA  
SPADLIEYSAPNPNDPEYSGOMGNSYIPNOITVSSOLDPNDPEYSGILDEITTA

```

3'UTR      QEQSNLSWFFENINGCASSSDSLNIGEE
1127.  .1423
/gene="MYB108"
BASE COUNT  498 a  273 c  282 g  370 t

```

Query Match	19.5%	Score 167.8;	DB 8;	Length 1423;
Best Local Similarity	69.28;	Pred. No. 1.7e-36;		
Matches 229; Conservative	0;	Mismatches 102;	Indels 0;	Gaps 0;

[illegible]

## RESULT 8

LOCUS	AP003140	133242 bp	DNA	linear	PLN 22-MAY-2003
DEFINITION	Oryza sativa genomic DNA, chromosome 1, BAC clone:OSJNBa0025p13.				
ACCESSION	AP003140				

VERSION	KEYWORDS	SOURCE ORGANISM
AP003140.2	GI:11164491	<i>Oryza sativa</i> (cultivar: Nipponbare) DNA, clone: OSJNBa0025p13. <i>Oryza sativa</i>

REFERENCE  
1 (sites)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

**AUTHORS** Sasaki, T., Matsumoto, T. and Yamamoto, K.  
**TITLE** *Oryza sativa* nipponbare(Az3) genomic DNA, chromosome 1, BAC clone:OSJNBa0025P13  
**JOURNAL** Published Only in Database (2001) In press

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (31-JAN-2001) Takuji Sasaki, Nat

Agrobiological Resources, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: [tsutsakiba@affrc.go.jp](mailto:tsutsakiba@affrc.go.jp), [URL: http://rpg.dna.affrc.go.jp/](http://rpg.dna.affrc.go.jp/),  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT  
On May 21, 2001 this sequence version replaced g1:12641874.  
Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLASTX2.0, BLASTY2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against

NCBI nonredundant protein database, nr (<http://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database dbEST. Protein homologs of the coding regions were searched against NCBI nonredundant protein database with BLAST2.0. ESTs containing

the identified cDNA sequences using BLASTN 2.0 with the corresponding DBJ accession no. and RCP clone ID. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level.

subdivided based on the protein name: the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program

unknown protein: a gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OSJNBa0025P13 clone has an overlap with

p0702012 (DBJ: AP002820) clone at the position 1 to 1534 of 5' end and with p0487002 (DBJ: AP002883) at the position 84175 to 133242 of 3' end. The sequence of this clone starts at the position 135799 of p0702012 and ends at the position 49068 of p0487002. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://trp.dna.afrc.go.jp/Genomeseq.html>.

## FEATURES

## SOURCE

```

Location/Qualifiers
1..133242
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="OSJNBa0025p13"
complement(join(4221..4246,4864..5040,5244..5909,
6141..6189))
/gene="OSJNBa0025p13.1"
complement(join(4221..4246,4864..5040,5244..5909,
6141..6189))
/gene="OSJNBa0025p13.1"
/codon_start=1
/product="putative protein kinase"
/db_xref="GI:14164492"
/translation="MYLPMKLVILFVLAALVYISICRICRRKKYKIHQAGDLOQENM
FSYLNFGGNAVEGITATEENFESEKTCIGAGYVSYYAALSKGRFAVKINATENG
RLINEQMFYEIEATWQIRHNITVKVEGCTARDFIVTKYKGGNLPAKLSYSTSA
SELDWKRRLCIADVAHALSYLHHDCSDPIVHRDVTTKNILLDEFRACISDFGKIL
LDADGSGHTRLAGTGYLAPELAYTTKTEKCVSYSGVVLELMLGSHGDFVSSIS
CPSKSTPMKDLDTLRPPAEVYALKIM"
join(9270..9380,9468..9560,9738..9797,9871..10143,
10245..10319,10435..10491,10609..10749,10829..10879,
10972..11067,11175..11258,11346..11448,11534..11616)
/gene="OSJNBa0025p13.2"
join(9270..9380,9468..9560,9738..9797,9871..10143,
10245..10319,10435..10491,10609..10749,10829..10879,
10972..11067,11175..11258,11346..11448,11534..11616)
/gene="OSJNBa0025p13.2"
/notes="contains ESTs C72171(E1142), AU093893(E1142)"
/codon_start=1
/product="putative shaggy-related protein kinase"
/protein_id="BAB55743.1"
/db_xref="GI:14164493"
/translation="MTSVGLVDSSSGFPESTSGADRLTDDISEMSIRKREAVVY
SGNMDIGHITVTVGNGSQPKOTISTIAERKAVGSGFVVAQALEGGEVAARK
VLQDARKKNEELQTMQVLDHPNVACLKHYFCSTAKEELVNLVLEVPETVARIRH
YNNKSQMPRLIVKLYMYQICRALYIHNCVYCHRDIKPONTLVNPHNQKLCDFG
SAKVLVGEENISYICSRYYRAPELIFGAREYTAIDVMSAGCVLAEILGQVDFGD
SGVDLVEITIKVLEGTPTREEIKHNPYTEFKFPQIKAPHMKIIFHKRMPSEAVDLYS
RILOYSPLHSCSALEVLHHPFDELRLPNRLNGRILPLPLNFKPRELKGASMEELY
KLVQNAKKKQCAPLGT"
join(15137..15820,15928..15963,16058..16849)
/gene="OSJNBa0025p13.3"
join(15137..15820,15928..15963,16058..16849)
/gene="OSJNBa0025p13.3"
/codon_start=1
/product="putative polygalacturonase-like protein"
/protein_id="BAB55744.1"
/db_xref="GI:14164494"
/translation="MELAAAGRTAAILALLALFASSFISAADGARSARHHNAKNAK
NAAPPSQAPGPBARHAPRAPHNGAPRPHGAPRPARAPRPPSDPPGAPGAPSA
PAGCAATVIVIDFGAAGDVTDDDTALCTAMPDTACADGAGCVLAAGARSLIHTT
VFTGPCGSGTLOYDGTIVAPSEPAITPANNKRMVLFYRADGVSILVAGCLIDGCKG
WMDLPLPKPHKGGNTHPCDSVPAMRFATISNNVTVRGKLVONSPEFHRFQNGCVRD
GLSDLSPLSPNTDGHVENTSDVLIINTVNSVDDCVSAGAGTLNVAHNEINCPGPH
GISISLGKAGTKACVANTVNNAVINHSNDGVSIGAGTIVGASVSAVAENRMAV
RNPILIDYYCTGKSCENETFAVENGVSYSGIRGTVDGPRPIHFGCSGAVCTNNT
LSVDLPLASGDTDDEPCNVYGNATTPVAVPVSCLSSGSPNTTEKKDQCY"
complement(join(21118..22066,22246..22397,23530..23691))
/gene="OSJNBa0025p13.4"
complement(join(21118..22066,22246..22397,23530..23691))
/gene="OSJNBa0025p13.4"
/notes="hypothetical protein"

```

```

/codon_start=1
/protein_id="BAB55745.1"
/db_xref="GI:14164495"
/translation="MRAIEQLRVLVAAAARAPAGRWTHRGARDYRRCGGQDRSARP
ATEAGGARVYTKNGTQTRSPYSGVPRKJDEKRRKTAAGHOSKEKRLCFRAIT
LGLRAKTPVHPHAPRTEVAASSTTSASAKKAVSSGRASRNANPISHLRLTSRO
ACILHRLPRLSRAGRCRHLRHVSYSRPTPPNNMALADTYCHALGKPSSEAMP
PTTVVTSFSPSPHTQLPRHLCHLRHLSHGCKSAVAADAVSHDSNR
HCKPAPPTQPPQRRRLALCRDRHGRKTISSAVIRSPMPASPLDPAVSALVTATSI
PAISNPADADASKSSMPARTPLSPSPPPPPSPSSAPATPNCRRHPTTFHPRR
HHHQLLPHTTQPPPAFPDPATAARI"
complement(join(27109..27305,28392..28419))
/gene="OSJNBa0025p13.5"
complement(join(27109..27305,28392..28419))
/gene="OSJNBa0025p13.5"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB55746.1"
/db_xref="GI:14164496"
/translation="MDVVIIGKTKKQTANQKKANAMHSQDITSSIGNFDIDKKK
DNLNRSYTVIAHFEYRSIVANRLALIK"
complement(join(31012..31020,31243..31294,31358..31478,
32907..33108,34245..34337))
/gene="OSJNBa0025p13.6"
complement(join(31012..31020,31243..31294,31358..31478,
32907..33108,34245..34337))
/gene="OSJNBa0025p13.6"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB55747.1"
/db_xref="GI:14164497"
/translation="MHERVILSEYNDKEITGFHNPOLHGMCKOKERTHPSLPYRRR
MPSDKFAENMRANOSLSPDVSHVHFVVLASVLAISGRKRSPTPDGKQVSR
TSVATSCSDCARLIPSVVNVILPLDLSGMBDRVCPPLFVPICGFKQKPPNVW"
41044..44517
/gene="OSJNBa0025p13.7"
41044..44517
/gene="OSJNBa0025p13.7"
/codon_start=1
/product="putative GAG-POL precursor"
/protein_id="BAB55748.1"
/db_xref="GI:14164498"
/translation="MPCVPRVEYIENHLAVPGARPPVROKVRQADPEROAFIREVART
LEAGFIREYIHEMLVNPYIVKRAKKLMKCIDTYDLMKACRQDYPLSRIDQYDST
AGDCLICFTSGTHQIMNAREDEKTAFTIPIGTCITTPFGIKRNPFTQTRR
ISGSQIGNVAAVYVDLVKTRNDQTLSDLAETFEESLRARIKILNPKCVFQVPAG
KILGFIYSARGIEANPEKIRAIERMPSPSKLDVQCVTCMAALSRFISRLGEKALPL
FKLKSGSFTTEEAERAFOLKAYLSPPVLAPEDEPLALLAAPQVNSALV
VERDEENPISANHRPASTRPGEGGEAPENGGPPRPAAGAPRPACTVAGADPO
DGGGAGAGRPRLSPDPEVNGTEAGCAPRGLSDEERPGDGAARGEPRRKYQRRVYF
VSEALDATTRIPQAKKMLIATLVASRKLRHFQANRYTVYTSYLGQILINRETCR
VVMATLESEDFLREPRHAISQALJTDVAEMTPAPEVVSAPLEASSGSQLPHATV
VMQFDSLSLQAGAGVLTLSPSGDVLRVLNLDPRATNNMVEGLGLAGLVNAGLG
IRLLVLAGDSOLVNVQVYKEYKCSDPOMAVYQVARMRHRFDGIELRHVPRDNAVA
DELSRLASSRAQTPGCAFEERLAOPRARDPLGENDAPRPRPIGVQVSGEGSAPS
SPRLIATWTEIOTIYADKTLPLDRGSESEVRISRYVVEGTLYRRAAGVLCIP
RBOGVLLADYHNEGEGASASRTLYGKAFROGFYVPTNLNANVULYRRCACQCHAR
QTHQPAQLOTTIPLSWPAVWGDLIDGPRRABGGEETQVAVADVETKPEATYVIKI
DKHSALKEIRGITARFQVNPRIITNNGTQETSELGCDKMDGILKCPASPPNSNG
OVERANAEFLIKGKTFTFNIILKHHGSMTEKLEPLAMARTTSPATGTSFFLVYGA
EVALPELTLRSRPTVCEADQDRLRDOLYLEERRRALLAARAYQOQSRLRHQR
HVAERSLACVDLILRVRYQTRAGLSKSPMBEPRYVIGVRRGVSRLATDGTDELNP
WNIETLRRYRP"
complement(join(45140..45474,45602..46322,46382..47023,
47154..48413))
/gene="OSJNBa0025p13.8"
complement(join(45140..45474,45602..46322,46382..47023,
47154..48413))
/gene="OSJNBa0025p13.8"
/notes="hypothetical protein
similar to Oryza sativa chromosome 3, OSJNBa0067E01.7"
/codon_start=1
/protein_id="BAB55749.1"

```



```

gene
CDS
    /protein_id="BAB67845.1"
    /db_xref="GI:15623785"
    /translation="MRTHTRTGGRAPGAPVCDTHDAASQVLMLEKLRDPAVNG
    ATRISPERNITKYSRPLRLICRGAPPTGRORORELPADGADPRVRGCEARR
    GTAGCCVSGDSRGTSVGVGARGARLRBRROMLBSGVCHCDAAACATGNNARPMR
    CGCCPAGGTSAVVPPIPRRRRMRMSTIVPPSNLINTGSSSSSHCKIKKEBETTP
    AAT"
    join(12412..12593,13512..13574,14331..14862,14954..15319)
    /gene="P0487H02.5"
    join(12412..12593,13512..13574,14331..14862,14954..15319)
    /gene="P0487H02.5"
    /note="contains EST C99297 (E10619)"
    /codon_start=1
    /product="putative MtN21"
    /protein_id="BAB67846.1"
    /db_xref="GI:15623786"
    /translation="MGRRESPLTAMVNOGLFAGMNVSKLADTGMSPVYLAIANI
    IAAVFLAPRAYFRERTKSGVITTKVIVQIFSSIFGATLNOVLFPVGLKSTPTVA
    CALSNTLPALTEMAAARFRESVRLSAAGAAQAFVGVGSMIMPYKQPLRLM
    ASDIHRFAESAASGAAPAGAAVAGVLIILSCAAMWVEILOTKMSERSAPYT
    STTIMCLMAGVQAGVSAAMDRAVAWKLGFDIRLSVLYIGVSGIAFPALSMCTQ
    VRGPLEVSMSPMLVVAIVGMAILDEKIHVGSALVIAGLVYMWLNGAREMS
    PSDLDGGGGGGVVELNGGADAATLIPVCTTNNKHTTRNCCSN"
    complement(join(21707..21839,22076..22099,22760..22917))
    /gene="P0487H02.6"
    complement(join(21707..21839,22076..22099,22760..22917))
    /gene="P0487H02.6"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB67847.1"
    /db_xref="GI:15623787"
    /translation="MVREGRDLGAGAGRAETGGEGVCAPTPHIGDMEETGECI
    VRADQRDRKMPVRCNGMRKTHARRREVSDSASACDDCRKSTEAKFFSLYLHLH
    PY"
    complement(join(23501..23554,23930..24028,24083..24274))
    /gene="P0487H02.7"
    complement(join(23501..23554,23930..24028,24083..24274))
    /gene="P0487H02.7"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB67848.1"
    /db_xref="GI:15623788"
    /translation="MMTECRRLFEENTSRVGAONDLTRAISSDGRVQMLEEVA
    RDLQKMEIIVAGNEQRLGLKMSFTSLQDLAECGVKSSIPAHPEETSEPKRGVQR
    RIGNNVMDDEM"
    complement(join(27604..27686,27736..28503,29633..29675))
    /gene="P0487H02.8"
    complement(join(27604..27686,27736..28503,29633..29675))
    /gene="P0487H02.8"
    /note="hypothetical protein
    similar to Oryza sativa chromosome 6, P0514G12.3"
    /codon_start=1
    /protein_id="BAB67849.1"
    /db_xref="GI:15623789"
    /translation="MKGASHTLTAACHGCTDSANQCCVSNTRVCRPRDGLGTPD
    LRVAGRLAKRMLQDSCHPHTSDLKASERSVDMFASTFSILGDQSTLFWTDK
    WIDGRSIASLAPDLHVVPREFGSRFVAAGLANNSWGDICGALTFVPIQFLWMD
    AVVPTQISPEDEHVLWMTGQDSAGSAYQAFPGHSGFCADLIMIAKSPAKKF
    FLWFAEORCMWADLLQKRGIDSHSACPQACQLEFANHILIDCVARQWMLPGSRIG
    GHPLEGVVSTFMAASTPMYS"
    complement(join(30213..30418,31118..31292,32407..32601,
    33634..33894,33932..34047,34149..34246,34509..34561))
    /gene="P0487H02.9"
    complement(join(30213..30418,31118..31292,32407..32601,
    33634..33894,33932..34047,34149..34246,34509..34561))
    /gene="P0487H02.9"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB67850.1"
    /db_xref="GI:15623790"
    /translation="MGRHVGVGLAMREGARPCPNPGRPPRRLLFPGRYPLRRGSR
    YPEVERDSEAGRRFRFRFSRSSRRHGRGAVAPAVAVVPLPSSSSSPAPVPS
    ATTDPGSGTCKKATPBLGLAAGSAATCKSSSSRAARRRRRVTTTLAPPAPEDESG

```

```

gene
CDS
    /protein_id="BAB67851.1"
    /db_xref="GI:15623791"
    /translation="MDAPCGWTTSAAAVAEEEDVRRPMTVEEDAVLAGVAAANG
    RNVELRAAGLKRRTGSCRLRLNLYLRPDRVGGDTTPOEOLLILELHRCMNRMSRIA
    OHMPGRDNEIKNMYRTVRKAROLGCDVNSROKRLMMPRIYERIHAAASS
    EBAAPPPCAAAARASHSCGHSCHSDPTSTSSNAGSSVTHGDEPPSTNNHLLTMAVYT
    AAADSSDECGSGSATSTSVGISTYMPFGSWSLLARAYDDGDADSDLLPPOFADTGT
    DMCWMSNLEDIWSQPY"
    complement(join(41992..42118,43485..43552))
    /gene="P0487H02.11"
    complement(join(41992..42118,43485..43552))
    /gene="P0487H02.11"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB67851.1"
    /db_xref="GI:15623791"
    /translation="MDAPCGWTTSAAAVAEEEDVRRPMTVEEDAVLAGVAAANG
    RNVELRAAGLKRRTGSCRLRLNLYLRPDRVGGDTTPOEOLLILELHRCMNRMSRIA
    OHMPGRDNEIKNMYRTVRKAROLGCDVNSROKRLMMPRIYERIHAAASS
    EBAAPPPCAAAARASHSCGHSCHSDPTSTSSNAGSSVTHGDEPPSTNNHLLTMAVYT
    AAADSSDECGSGSATSTSVGISTYMPFGSWSLLARAYDDGDADSDLLPPOFADTGT
    DMCWMSNLEDIWSQPY"
    complement(join(41992..42118,43485..43552))
    /gene="P0487H02.11"
    complement(join(41992..42118,43485..43552))
    /gene="P0487H02.11"
    /note="hypothetical protein"
    /codon_start=1

Query Match      19.3%  Score 166.4; DB 8; Length 145491;
Best local Similarity 68.5%; Pred. No. 1e-35;
Matches 220; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 63 gtcccaagaatgctgagatgagaaaggccgtgagcaatgagaaagatctcatc 122
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35950 GCGGAGAGAGAGAGATGTCGAGAGAGCGCGTGGAGCGAGAGAGAGAGAGCGGTCTGCG 36009

QY 123 caactacatagcgaatcaagaggagagcagcttggaactccctagccaagtcgctgct 182
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36010 CGGCTACGTGCGCCGACGCGAGGCGGCGGTGAACGAGCTGCGCGCGCGCGGCGCT 36069

QY 183 aaaaagctaccggaagagatgctgctcggctgagcaatctcgcacccgagctcg 242
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36070 GACGGCAGCGGAGAGAGAGCGTCCGCGTCCGCTCACTCACTCACTCACTCACTCACT 36129

QY 243 gagaagcaacatactactcgaagagcagctcctgatactcgaactgcatgccaagtggag 302
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36130 CCGCGGCGACTTCACCCGAGGAGCAGCTGCTCACTCACTCACTCACTCACTCACTCACT 36189

QY 303 aaacaggtgctcaaaattgcaagatctcccggaagagcagcagaatgataaaga 362
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36190 GAACCCCTGCTCCAGCATCGCGACGACATCGCGGAGAGAGAGAGAGAGAGATCAGAA 36249

QY 363 ctctgagagactagaatccaaagacatcaagca 398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36250 CTACTGGCGGAGACGAGGTGACAGAAAGCAGCAAGCA 36285

RESULT 10
AF334815 1140 bp mRNA linear PLN 30-AUG-2001
LOCUS Arabidopsis thaliana putative transcription factor (MYB116) mRNA,
DEFINITION complete cds.
ACCESSION AF334815
VERSION AF334815.2 GI:15375294
KEYWORDS
SOURCE
ORGANISM
    thale cress.
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
    1 (bases 1 to 1140)
    Stracke,R., Werber,M. and Weishaar,B.
    The R2R3-MYB gene family in Arabidopsis thaliana
    Curr. Opin. Plant Biol. 4 (5), 447-456 (2001)
MEDLINE 21481677

```



REFERENCE	AUTHORS	TITLE	JOURNAL
PUBMED 11597504	2 (bases 1 to 1140)	Stracke,R. and Weisshaar,B.	Direct Submission
REFERENCE	Submitted (10-JAN-2001)	Dept. Plant Breeding and Yield Physiology,	Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg
AUTHORS	10. Koeln D-50829, Germany		
TITLE	3 (bases 1 to 1140)		
JOURNAL	Stracke,R. and Weisshaar,B.		
REFERENCE	Submitted (29-AUG-2001)	Dept. Plant Breeding and Yield Physiology,	Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg
AUTHORS	10. Koeln D-50829, Germany		
TITLE	Sequence update by submitter		
JOURNAL	On Aug 30, 2001 this sequence version replaced gi:13430155.		
REMARK			
COMMENT			
FEATURES			
source	1..1140	/organism="Arabidopsis thaliana"	
gene	/db_xref="taxon:3702"	/note="ecotype: Columbia"	
5'UTR	1..1140	/gene="MYB116"	
CDS	1..129	/gene="MYB116"	
	130..981	/note="At1g25340; R2R3-MYB family member"	
	/codon_start=1	/product="putative transcription factor MYB116"	
	/protein_id="AAK5748.2"		
	/db_xref="GI:15375295"		
	/translation="MSNITKKKCGNEGAERQKRPWTLIEDTLITNYISHGGRWNN		
	LLAASGLIKRAGKSCRLRLMLNYLKPDIKRGILTPQEDLLILELSKGNRMSKISKYU		
	PGFPRNDIKWYRFRVQKQAROLINDSNHKEFIVSFWPRLINELKONSTYNNIK		
	ANAPDILGPIRLRDSKDLGFNNMDCSTMSDEDLKTSSQPMDSDELFTTMSLEGSGGSS		
	OCSEVYVSFPCELEEEYVAVAVMGSSDISLALHDCVAVDSKIEDVYTQDLMMNDDIMQER		
	NEVIAHFN		
	982..1140	/gene="MYB116"	
BASE COUNT	397 a 202 c 222 g 319 t		
ORIGIN			
Query Match	18.8%; Score 162.2; DB 8; Length 1140;		
Best Local Similarity	67.8%; Pred. No. 6,2e-35;		
Matches 227; Conservative	0; Mismatches 108; Indels 0; Gaps 0;		
55	agtgtaagtcccaagatgctcgaggtgagaaaggccgtgacgaatgaaaggatctc	114	
157	AATGGAATGAAGAGCGGTGACAGACAGAGAGAAAGGCGCTTGACACACTCGAGAAACACT	216	
115	atcccatcaactatcatctgcaatctacggcggaagcgagttggaactccctagccaaagt	174	
217	CTTTCACCAATTAATCATTTCCATTAACGGGTGAAGGCCGATGGAATCTGCTCGCTAAATCT	276	
175	gctggtctcaaaacgtacacggaagaggtgtgctgcgcgtgacgtgacatctatctgcaacc	234	
277	TCTGGGCTTAAAGACGACGAGAAAGTGTGTACATTGAGAGTGGTTGATTAATCTTAAACC	336	
235	gacgtccggaagcgcaacatctactctgaagagcagctctgcatcatgaaactgacgc	294	
337	GACATTAAGCGGTGGAATCTCACTCCTCAAGAACCACTTTTAATCCTTGAGCTCCATCTCT	396	
295	aagttgggaagaacaggtgtgtctaaactgtcaaaagctctctccggaaagagactgaacatgag	354	
397	AAATGGGGTAATAGGTGGTCAAAATTTTCGAAGTATTTCACGAGGAAGAACACAGACAAACAT	456	
355	ataaagaactcttgagagactagaatccaaagca	389	
457	ATCAAAACTACTGGAGAACTGAGGTCCGAAACA	491	

[illegible]

Oy	205	cgcgcccgagtgagctgaaactatctgcagaccgaagcgcggagaaagcaacatcactacgag	264
Db	303	AGATTGCGGGTGCTAAATTATCTCCGGCCGGATATCCGCCGTGGAGACATATCCCTTGAA	362
Oy	265	gagcagccctcgatcactggaactgcabgcgcaagtcggggaaacagctgctctaaattgca	324
Db	363	GACACATTTATCATCTCTTGAACTCCATTTCTGGTTGGGGAATTCGGTGGTCAAAATTTGCT	422
Oy	325	aagcactctcccggaagagactgcacabtgaglaaagaactctgagagactagatcaca	384
Db	423	CACACTTTTACCGGGAAGAACATTAACGATTAAGAATTTATTGGAAACACCGGTTCAA	482
Oy	385	aagca	389
Db	483	AAGCA	487

RESULT	12				
AF034133					
LOCUS					
DEFINITION	AF034133	1033 bp	mRNA	linear	PLN 02-MAR-1998
ACCESSION	Gossypium hirsutum MYB-like DNA-binding domain protein (Cmy-N)				
VERSION	AF034133				
KEYWORDS	mRNA, complete cds.				
SOURCE	AF034133.1	GI:2921337			
ORGANISM	upland cotton.				
	Gossypium hirsutum				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
REFERENCE	1 (bases 1 to 1033)				
AUTHORS	Loguercio, L.L., Zhang, J., and Wilkins, T.A.				
TITLE	Structure and expression of six classes of myb-domain genes in				
	allootetraploid cotton ( <i>Gossypium hirsutum</i> L.)				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1033)				
AUTHORS	Loguercio, L.L., Zhang, J., and Wilkins, T.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-NOV-1997) Agronomy & Range Science, University of				
	California, One Shields Ave., Davis, CA 95616-8515, USA				
FEATURES	Location/Qualifiers				
source	1..1033				

gene	repeat_region	repeat_region	BASE COUNT	308 a	223 c	253 g	249 t
gene	1.1033	/dev_stage="3 days pre-anthesis"					
gene	106.987	/gene="Cmy-N"					
gene	106.987	/note="MYB-domain gene N"					
gene	106.987	/gene="Cmy-N"					
gene	106.987	/function="putative MYB-like transcription factor"					
gene	106.987	/note="similar to MYB A encoded by GenBank Accession Number L04497"					
gene	106.987	/codon_start="1"					
gene	106.987	/product="MYB-like DNA-binding domain protein"					
gene	106.987	/protein_id="AAC004719.1"					
gene	106.987	/db_xref="GI:2921338"					
gene	106.987	/translation="MPTREVSYLKIQEYVKKRWRLKGPWTEEDSMRLRAYVNIHGEGRWNAVARLSGLRRTGSKSRRLWMLNPAEIKGNISLEQLLLELHSHWGNWSKIAIAHLPGRTDNEIKMYKRTFVKOKACVNSQVDFADARYAVIPLVIRTRASSPSPSPTTNTYNDRIISNTSSSSQMSAYNAGSQVNDPSLIPELSGTSSDLSITQVSSVDLTIDYINPQSLNSNTHKKGGLGEKEGAAATWGRDEEPQATEEHSNGMLVGGESSMDTVNEENWFLQOQDHHDDGI"					
gene	106.987	163.318					
gene	106.987	/note="putative MYB DNA-binding domain repeat R2"					
gene	106.987	319.474					
gene	106.987	/note="putative MYB DNA-binding domain repeat R3"					

Query Match	17.6%	Score 151.4;	DB 8;	Length 1033;
Best Local Similarity	67.0%	Pred. No. 7.1e32;		
Matches 215; Conservative	0;	Mismatches 106;	Indels 0;	Gaps 0;

QY	78	ggcgcgaataaagggccgcgcygaacgatcggaaagagatcctacatccatcacactatagcgaa	137
DB	162	gcttaagcgaagagctccatcgactgtaagaagagacactccatccttaagccgctatctcaatat	221
QY	138	tcaacggcgaagcgacgtctggaactccctcagcccaaacgcgcgcgtctctaaacgtaacggaa	197
DB	222	ccacgcccaagctccctctgaaacgctgttgctcgttattcagagattgaaagaaacccgtaa	281
QY	198	gaattctcgcgtccgcgctgacgaactatcgcgaacccgacgctcccgagaggaacaacac	257
DB	282	aagctgcagcagattgagatcgctgacactatctgcacacagaanaatcaaacagagaacaatcag	341
QY	258	tactgagagcagcgtccctgatacatalggaactgcatalgccaagtcgggaaacaagctgcttaa	317
DB	342	ccttcgaagacagcagattatgattcttgaaactcattctcgctgggggaaacaggtgctcgaa	401
QY	318	aattgcgaagcactctctcccgaaaggaactgacaatgagataaagaactctcgagggactag	377
DB	402	aattgcacaacactctgctcggaacacagacaataagaataaagcattatttgaggaacacag	461
QY	378	aatccaaagcacatcgaagca	358
DB	462	agttccgaagcagggcccaagca	482

RESULT	13		
AF427146			
LOCUS		1513 bp	mRNA
DEFINITION	2ea mays myb-like transcription factor (P)	linear	PLN 29-OCT-2001
ACCESSION	AF427146		
VERSION	AF427146.1		
	GI:16507119		

ORGANISM	Zea mays.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Euphyllophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 1513)
AUTHORS	Hartmann,U., Sagasser,M. and Weisshaar,B.
TITLE	Separation of cis-acting elements required for light-responsive and tissue-specific expression of co-regulated phenylpropanoid biosynthesis gene
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1513)
AUTHORS	Hartmann,U., Sagasser,M. and Weisshaar,B.
TITLE	Direct Submission
JOURNAL	Submitted (01-Oct-2001) Salamini, Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, NRW 50829, Germany
FEATURES	Location/Qualifiers
source	1..1513

```

gene
1..1513
/feature="P"
179..1186
/gene="P"
/feature="DNA-binding domain"
/codon_start=1
/product="myb-like transcription factor"
/protein_id="A124047.1"
/db_xref="GI:16507120."
/translation="MGRTPCCEKVKLRGMTAEEDQLLVYIAEHGEGSMSRLPRNLTGLRGCSCRLIMWYLRADVYKGNISKEEDIILKLTATGNRSLIASHLPGRTDDEIKYNWNSHLSRQIHITTKRYKTAGDDPAIALIMSKLOSADRRRCGRIPGRPKASAEKTKQADADQPCGEAKGPAALASPRHSUVVNGEPPQNPSSSGSTTAEEGTSDDAAGPWTELEIGLDVWGEADSEMDALMTGPGGHSAALGAGVCEAOVDDLPEMDM

```









OX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Bann J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shino P., Yamada K., Shinozaki K.,  
 RA Becker J., Theologis A., Davis R.W.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF042894; AAK6834.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 293 AA; 32221 MW; A12532EAE45795A CRC64;

Query Match 38.5%; Score 574.5; DB 10; Length 293;  
 Best Local Similarity 47.0%; Pred. No. 4e-43;  
 Matches 133; Conservative 17; Mismatches 62; Indels 71; Gaps 6;

OY 5 AETRYDSFETSGSSVDLGMALPGLAPNLSSASVSASASASAKKIRKPYTTIKSRSS 64  
 DB 14 AETSDATETT-----IATTEGAPPEKKVKKATYTTIKSRSS 50  
 OY 65 WSGQENDKFLKALQFLDRDKKIEAFVSGKTVIOIRSHAQKFLVKNGKTRHVPVPPR 124  
 DB 51 WTEGHDKFLKALQFLDRDKKIEAFVSGKTVIOIRSHAQKFLVKNGKTRHVPVPPR 110  
 OY 125 KRASHPYOKASKNPVSGQVSTAPPTAATQDSCGYPRASSSILTKSGSSCPTVSSW 184  
 DB 111 KRKAHPYOKASKNPVSGQVSTAPPTAATQDSCGYPRASSSILTKSGSSCPTVSSW 164  
 OY 185 VHHITPIDASFEVKDDGCPGIEGNCSS-----GSTSSPPTWPCSEIPEKVK---- 236  
 DB 165 I-----PPEDELDTLCGAEVDVGSNDISSETSSASGICGSSRLSD 206  
 OY 237 -----PDFSQYKFTGVSVPDPTDHLKKLKE 263  
 DB 207 TKGRLAKQAPSMHGLPDAFVYNGFISGVDPDPSKGRMKLKE 249

RESULT 5  
 O91Z01 PRELIMINARY; PRT; 307 AA.  
 AC O91Z01;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE HYPOHETICAL 33.9 KDA PROTEIN.  
 GN F9G14.150.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Ardiles M., Buyschaert C., Dasseville R.,  
 RA De Clerck R., De Keyser A., Neyt P., Kouze P., Van den Daele H.,  
 RA Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,  
 RA Rudd S., Lemcke K., Mayer K.F.X.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL162973; CAB86038.1; -;  
 DR InterPro: IPR001005; MYB\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 1.  
 DR SMART: SM00395; SANT; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 307 AA; 33896 MW; 0F1C6D375A62B050 CRC64;

Query Match 38.0%; Score 567.5; DB 10; Length 307;  
 Best Local Similarity 44.8%; Pred. No. 1.8e-42;  
 Matches 133; Conservative 17; Mismatches 62; Indels 85; Gaps 6;

OY 5 AETRYDSFETSGSSVDLGMALPGLAPNLSSASVSASASAKKIRKPYTTIKSRSS 64  
 DB 14 AETSDATETT-----IATTEGAPPEKKVKKATYTTIKSRSS 50  
 OY 65 WSGQENDKFLKALQFLDRDKKIEAFVSGKTVIOIRSHAQKFLVKNGKTRHVPVPPR 124  
 DB 51 WTEGHDKFLKALQFLDRDKKIEAFVSGKTVIOIRSHAQKFLVKNGKTRHVPVPPR 110  
 OY 125 KRASHPYOKASKNPVSGQVSTAPPTAATQDSCGYPRASSSILTKSGSSCPTVSSW 184  
 DB 111 KRKAHPYOKASKNPVSGQVSTAPPTAATQDSCGYPRASSSILTKSGSSCPTVSSW 164  
 OY 185 VHHITPIDASFEVKDDGCPGIEGNCSS-----GSTSSPPTWPCSEIPEKVK---- 236  
 DB 165 I-----PPEDELDTLCGAEVDVGSNDISSETSSASGICGSSRLSD 206  
 OY 227 PCSEIPEKVK-----PDFSQYKFTGVSVPDPTDHLKKLKE 263  
 DB 207 SASGIGSSRLSDSKGLRLAKQAPSMHGLPDAFVYNGFISGVDPDPSKGRMKLKE 263

RESULT 6  
 O04605 PRELIMINARY; PRT; 213 AA.  
 AC O04605;  
 DT 01-JUL-1997 (Tremblrel. 04, Created)  
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE A\_I6002N01.20 PROTEIN.  
 GN A\_I6002N01.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Scheet P., Maggi L.;  
 RT "The sequence of A. thaliana IG002N01."  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Wash-U;  
 RT "The A. thaliana Genome Sequencing Project";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Waterston R.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF007269; AAB61027.1; -;  
 DR InterPro: IPR001005; MYB\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 1.  
 DR SMART: SM00395; SANT; 1.  
 DR PROSITE: PS00900; MYR.3; 1.  
 SQ SEQUENCE 213 AA; 23616 MW; 8C76CCEB9AB87A68 CRC64;

Query Match 33.7%; Score 502.5; DB 10; Length 213;  
 Best Local Similarity 56.4%; Pred. No. 6.7e-37;  
 Matches 106; Conservative 20; Mismatches 45; Indels 17; Gaps 5;

OY 22 DLVGMALP-----GLAPNLSSASVSASASAKKIRKPYTTIKSRSSDQEHKFLKFA 76  
 DB 15 DSSNMSLPSSDGFSGIPATGRFS-TVFSFSDPTTKIRKPYTTIKSRREMTDDEHDKFLKFA 73  
 OY 77 LQFLDRDKKIEAFVSGKTVIOIRSHAQKFLVKNGKTRHVPVPPRKRKASHPYPOKA 136

```

Db 74 LHLFDRMKKIEAFVGSKTIVQIRSHAQKFLKVOKSGANEHLPPPKRKASHPYPIKA 133
QY 137 SKNVYSQGVSTAFPTAAT--QLDGGYYPRAESSILTKSGSCPIVSSVNHHTI-----P 190
Db 134 PKNV-----AVTSLPSSSTLPLEPGVLVSSDSKSLMGNOAVCASTSSSMNHSTNLPKP 188
QY 191 SIDSAYE 198
Db 189 VIEGSEFD 196

RESULT 7
Q9SR43 PRELIMINARY: PRT; 125 AA.
AC Q9SF43;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE PUTATIVE MYB-RELATED PROTEIN.
GN F11F8_19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxId=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maitl R., Rensing C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 3 BAC F11F8 genomic sequence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC016661; AAF23291.1;
DR InterPro: IPR001005; MYB-DNA-bind.
DR Pfam: PF00249; myb-DNA-binding; 1.
DR SMART: SM00395; SANT; 1.
DR PROSITE: PS50090; MYB_3; 1.
SQ SEQUENCE 125 AA: 14063 MW: 8A92A097F3F726EB CRC64;

Query Match 31.1%; Score 463.5; DB 10; Length 125;
Best Local Similarity 83.3%; Pred. No. 9.8e-34;
Matches 90; Conservative 5; Mismatches 12; Indels 1; Gaps 1;

QY 32 APNLSSASVSASSEDSSAKKIRKPYTTTKSRESWSEDEHDKFLDALQLFDRMKKIEAFV 91
Db 14 APPRPSTDAVA-EGSSKVKRRPYTTTKRESSETEDEHDKFLDALQLFDRMKKIEDV 72
QY 92 GSKTVIOIRSHAQKFLKVOKNGTREVPPRRKRASHPYPOKASKN 139
Db 73 GSKTVIOIRSHAQKFLKVOKNGTLAHVPPRRKRASHPYPOKASKN 120

RESULT 8
Q23701 PRELIMINARY: PRT; 162 AA.
AC Q23701;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MYB-RELATED PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxId=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINP=98332714; PubMed=9666060;
RA Terry N., Gleason J., De Keyser A., Van Den Daele H., Ardiles W.,

```

```

RA Neyt P., De Clercq R., Coppieters J., Delais P., Villarroel R.,
RA Rouze P., Van Montagu M.;
RT "Sequence analysis of a 40-kb Arabidopsis thaliana genomic region
RT located at the top of chromosome I.";
RL Gene 215:11-17(1998).
DR EMBL: Y12776; CAA73305.1;
DR InterPro: IPR001003; MYB-DNA-bind.
DR Pfam: PF00249; myb-DNA-binding; 1.
DR SMART: SM00395; SANT; 1.
DR PROSITE: PS50090; MYB_3; 1.
SQ SEQUENCE 162 AA: 18929 MW: 34D5D44F1D796FAF CRC64;

Query Match 30.9%; Score 460.5; DB 10; Length 162;
Best Local Similarity 73.0%; Pred. No. 2.6e-33;
Matches 89; Conservative 12; Mismatches 12; Indels 9; Gaps 2;

QY 26 MALPG---LAPNLSSASVSASAS-----EDSAKKIRKPYTTTKSRESWSEDEHDKFL 76
Db 16 MSLPGFNLPHTATTPVSIKSNRTMSFEEDPTKVKRKPYYTTTKSREWTEDEHDKFL 75
QY 77 LQLFDRMKKIEAFVGSKTIVQIRSHAQKFLKVOKNGTREVPPRRKRASHPYPOKA 136
Db 76 LHLFDRMKKIEAFVGSKTIVQIRSHAQKFLKVOKNGTKEHLPPPKRKANHPPYPOKA 135
QY 137 SK 138
Db 136 PK 137

RESULT 9
Q943P3 PRELIMINARY: PRT; 255 AA.
AC Q943P3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE P0011G08.21 PROTEIN.
GN P0011G08.21
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
ON NCBI_TaxId=4530;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone: P0011G08."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003225; BAB64649.1;
SQ SEQUENCE 255 AA: 28469 MW: B1B946F8E342A67E CRC64;

Query Match 26.2%; Score 391; DB 10; Length 255;
Best Local Similarity 39.5%; Pred. No. 7.1e-27;
Matches 98; Conservative 29; Mismatches 81; Indels 40; Gaps 5;

QY 34 NLSSASVSASSEDSSAKKIRKPYTTTKSRESWSEDEHDKFLDALQLFDRMKKIEAFVGS 93
Db 19 NTNNSNSSSSEMPGKRRKRPYYTTTKRERWSEDEHDKFLDALQIMYGDMKKIEHNGT 78
QY 94 KTVIOIRSHAQKFLKVOKNGTREVPPRRKRASHPYPOKASKNVYSQGVSTAFPTA 153
Db 79 KTTIOIRSHAQKFLKVOKMGIAAGLPQYPRRLVMQOOQSSPVAASVAATRAIILHG 138
QY 154 ATQLDSGYYPRAESSILTKS-----GSSCPVSSVNHHTIPISIDASEVEKDDGPGPIET 209
Db 139 OPOCLPRHHNVAVOSSICGSSFMGAASFNTSMDTGTSTSEMATASIYODE----- 189
QY 210 GNNCSGSTRSSPPMPPESEIPEKVKPD---FSQYVKRTIGVPPDPST---TDHLKTLK 262

```







DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)  
 DT 01-MAY-2000 (TReMBLrel. 13, last annotation update)  
 DE MYB-RELATED PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. IR-BB21;  
 RA Han F., Kilian A., Chen J.P., Kudrna D., Steffenson B., Yamamoto K.,  
 RA Matsumoto T., Sasaki T., Kleinhoofs A.;  
 RT \*Sequence analysis of a rice BAC covering the syntenous barley Rpg1  
 RT region.\*;  
 RL Genome 0:0-0(1999).  
 DR EMBL; AF149806; AAF00137.1; -.  
 SQ SEQUENCE 103 AA: 11895 MW: 96535D5F069B32EA CRC64;

Query Match 11.5%; Score 171.5; DB 10; Length 103;  
 Best Local Similarity 37.9%; Pred. No. 7.4e-08;  
 Matches 39; Conservative 12; Mismatches 19; Indels 33; Gaps 3;  
 OY 23 LVGMALPGGLAPNLSASVSASASSEDsAKKIRKPYTITKRSRSWSEQEHDK----- 72  
 DB 8 LIDSSSPGPGP-----GPEVEDGGGRKRVKPPYITKRSRESWTDPGQGVPMWRCSCKH 60  
 OY 73 -----FLEALQL-----FDRDMKKIEAFVGSKTVIQI 99  
 DB 61 PCAFFMWVPSYLFFAFFTCFYNNRDRDMKKIEAFVGSKTVIQV 103

Search completed: August 12, 2002, 10:11:40  
 Job time: 223 sec

1.

2.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:08:47 ; Search time 11.64 Seconds  
(without alignments)  
938.051 Million cell updates/sec

Title: US-09-640-211A-1076  
Sequence: 1 MPMLAETRYRDSFETSGSS.....EWIQLIKLCCTHPEPHNL 282

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SwissProt\_40:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	8.8	2472	1	Q9WU42 mus musculu
2	119	8.0	2517	1	NCR2_HUMAN
3	116.5	7.8	1733	1	Q9Y618 h nuclear r
4	105	7.0	1487	1	P33485 pseudorabie
5	104.5	7.0	919	1	Q9V618 dirosophila
6	102.5	6.9	562	1	O9V6H5 homo sapien
7	101	6.8	692	1	Q02457 saccharomyc
8	99.5	6.7	705	1	P79944 xenopus lae
9	98.5	6.6	939	1	P58462 mus musculu
10	98.5	6.6	942	1	Q03497 saccharomyc
11	97.5	6.5	1593	1	P54786 saccharomyc
12	96.5	6.5	444	1	P58397 homo sapien
13	94.5	6.3	977	1	Q12847 homo sapien
14	94	6.3	437	1	P97838 raltus norv
15	93	6.2	338	1	P26630 brachydanio
16	93	6.2	707	1	O90267 brachydanio
17	93	6.2	1273	1	P23246 homo sapien
18	93	6.2	1745	1	P38968 saccharomyc
19	92.5	6.2	437	1	P39447 mus musculu
20	92.5	6.2	568	1	O73917 oryzias lat
21	92.5	6.2	640	1	P42568 homo sapien
22	92.5	6.2	825	1	P46200 homo tauris
23	92.5	6.2	859	1	P32591 saccharomyc
24	92.5	6.2	872	1	O10362 schizosacch
25	92	6.2	677	1	O25434 mytilus cor
26	92	6.2	2363	1	Q9H334 homo sapien
27	92	6.2	2453	1	SPCO_MOUSE
28	91.5	6.1	736	1	Q62661 mus musculu
29	91.5	6.1	1005	1	O14641 homo sapien
30	91.5	6.1	1493	1	O13625 homo sapien
31	91	6.1	668	1	O06292 raltus norv
32	91	6.1	708	1	Q09761 schizosacch
33	91	6.1	940	1	P36417 dictyostell
					MA24_SCHCO
					P37938 schizophyll

34	91	6.1	1210	1	PER3_HUMAN	P56645 homo sapien
35	90.5	6.1	410	1	R1D_DROME	Q24106 dirosophila
36	90.5	6.1	600	1	HL2_MOUSE	O9WTV7 mus musculu
37	90.5	6.1	1505	1	CUT2_HUMAN	O14529 homo sapien
38	90.5	6.1	3924	1	ANK2_HUMAN	O01464 homo sapien
39	90	6.0	285	1	COAT_BPV4	O04755 lactococcus
40	90	6.0	951	1	SFR8_HUMAN	O12872 homo sapien
41	90	6.0	1185	1	DRPL_HUMAN	P54259 homo sapien
42	89.5	6.0	531	1	SIS2_CANTR	O12600 candida tro
43	89.5	6.0	551	1	GFY3_YEAST	P42294 saccharomyc
44	89.5	6.0	680	1	NASP_RABIT	P27123 oryctolagus
45	89	6.0	527	1	ABP2_SCHPO	O9USY4 schizosacch

## ALIGNMENTS

RESULT 1	ID	NCR2_MOUSE	STANDARD:	PRT:	2472 AA.
AC	Q9WU42; Q9WU43; Q9WU41;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Nuclear receptor co-repressor 2 (N-COR2) (silencing mediator of				
DE	retinoid acid and thyroid hormone receptor) (SMRT) (Thyroid-,				
DE	retinoid-acid-receptor-associated co-repressor) (TR3 receptor-				
DE	associating factor) (TRAC).				
GN	NCOR2 OR SMRT.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid-10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).				
RC	TISSUE-Spleen, and Brain;				
RX	MEDLINE-99178941; PubMed-10077563;				
RA	Ordenlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;				
RT	"Unique forms of human and mouse nuclear receptor corepressor SMRT.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA).				
RC	TISSUE-Embryo;				
RX	MEDLINE-99199215; PubMed-10097068;				
RA	Part E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;				
RT	"SMRT, a silencing mediator for retinoid and thyroid hormone				
RT	receptors-extended isoform that is more related to the nuclear				
RT	receptor corepressor.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).				
CC	- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME				
CC	NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS				
CC	PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.				
CC	- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B				
CC	AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES				
CC	WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE				
CC	ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.				
CC	- SUBCELLULAR LOCATION: Nuclear.				
CC	- ALTERNATIVE PRODUCTS: 2 ISOFORMS, ALPHA (SHOWN HERE) AND BETA; ARE				
CC	PRODUCED BY ALTERNATIVE SPLICING.				
CC	- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY				
CC	EMBRYOS.				
CC	- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT				
CC	ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2				
CC	AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-				
CC	INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION				
CC	DOMAINS (ID1 AND ID2).				
CC	- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED				
CC	SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND				
CC	SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES				
CC	FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR				
CC	SPECIFICITY.				
CC	- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.				
CC	- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.				

CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF113001; AAD20944.1; -  
 CC EMBL: AF113002; AAD20945.1; -  
 CC EMBL: AF125671; AAD22972.1; -  
 CC MGD: MGI:1337080; Nc012.  
 CC InterPro: IPR001005; MYB\_DNA\_bind.  
 CC Pfam: PF00249; myb\_DNA-binding; 2.  
 CC SMART: SM00395; SANT; 2.  
 CC PROSITE: PS50090; MYB\_3; 1.  
 CC Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 CC Coiled coil; Alternative splicing  
 CC -----  
 CC DOMAIN 155 207 COILED COIL (POTENTIAL).  
 CC 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).  
 CC DOMAIN 429 474 SANT-A (POTENTIAL).  
 CC DNA\_BIND 609 653 MYB.  
 CC DOMAIN 492 560 COILED COIL (POTENTIAL).  
 CC DOMAIN 652 682 COILED COIL (POTENTIAL).  
 CC DOMAIN 775 804 PRO-RICH.  
 CC DOMAIN 989 999 PRO-RICH.  
 CC 1351 1357 CORNR BOX OF 101.  
 CC 2094 2098 CORNR BOX OF 102.  
 CC 2296 2300 POLY-GLN.  
 CC 494 507 POLY-GLN.  
 CC 1615 1619 POLY-PRO.  
 CC 2434 2437 MISSING (IN ISOFORM BETA).  
 CC 36 254 M -> RL (IN REF. 2).  
 CC 176 176 PMLYDA -> RHVVR (IN REF. 2).  
 CC 366 402 D -> H (IN REF. 1; AAD20944).  
 CC 555 555 T -> A (IN REF. 1; AAD20944).  
 CC 756 785 V -> A (IN REF. 2).  
 CC 806 HHLPHRLMTRMNKKPRLQLPRQMPRSRLRRRSME  
 CC -> PSPAAPPATVVDKDECAAPAAPQTEDEAKSEAEI  
 CC DVG (IN REF. 2).  
 CC -----  
 CC 856 856 E -> K (IN REF. 1; AAD20945).  
 CC 859 859 E -> K (IN REF. 1; AAD20945).  
 CC 867 867 E -> K (IN REF. 1; AAD20945).  
 CC 895 895 E -> K (IN REF. 1; AAD20945).  
 CC 916 916 S -> F (IN REF. 1; AAD20944).  
 CC 975 975 I -> IO (IN REF. 1; AAD20944).  
 CC 1046 PKLPTEPRMSSGLPRPI -> QSYRLSPHACHRLPSH  
 CC (IN REF. 2).  
 CC PHADPSA -> TRAPPL (IN REF. 2).  
 CC 1073 1080 MISSING (IN REF. 2).  
 CC 1133 1133 MISSING (IN REF. 2).  
 CC 1149 1149 MISSING (IN REF. 2).  
 CC 1157 1157 G -> E (IN REF. 2).  
 CC 1172 1201 GSATSGSLKGLPSTAAAGPSTGSGTG -> APPVEA  
 CC SPRASQPGCRPOQLRHYPR (IN REF. 2).  
 CC -----  
 CC 1696 1696 A -> S (IN REF. 2).  
 CC 1855 1857 MISSING (IN REF. 2).  
 CC 1909 1909 A -> P (IN REF. 2).  
 CC 1913 1913 A -> G (IN REF. 2).  
 CC 1923 1923 G -> A (IN REF. 2).  
 CC 1923 1923 N -> S (IN REF. 2).  
 CC 1956 1956 A -> G (IN REF. 2).  
 CC 1968 1968 N -> S (IN REF. 2).  
 CC 2195 2196 TA -> AV (IN REF. 2).  
 CC 2214 LE -> SK (IN REF. 2).  
 CC 2224 T -> A (IN REF. 2).  
 CC 2472 AA; 270856 MW; 2A58FADF7B72858 CNG64;  
 CC SEQUENCE

Query Match 8.8%; Score 131; DB 1; Length 2472;

Best Local Similarity 23.0%; Pred. No. 0.15;  
 Matches 65; Conservative 34; Mismatches 141; Indels 42; Gaps 10;  
 QY 6 ETVRDSFETTSQSSVDYVGLPGLAPNLSASVSASASDESKKTRKRYTTITKSES- 64  
 Db 540 ELSEKEDTDDTSGENDENDEEAVASGRRTANSQGRKRGRITRSMANHEETAPQOSSE 599  
 QY 65 -----WSFOEDKPLEALQLFDROWKRTAEAVGSKTYIQRSHAKYFLKQKN 113  
 Db 600 LASMEMNSSWMTBEMETAKKGLLEHGRNNSAJARVGSNTV---SCKNFYFNKKR 655  
 QY 114 GTRHVPPEPRKKRASHDYPQKASKNVPVSOQVSTAPPTAATQD---SGYPR---AE 166  
 Db 656 QNLDDEILLOHKLKMEKERNAKRKRKTPAAASETAPPAEDEDEMAAGASANEELAE 715  
 QY 167 SSSILTKSGSSCPYV-----SSWVHTTIPSIDASFVEKDDGPGPIETGNMCSG 216  
 Db 716 EFAASQASGNEVPVAGDESGPAVANNSDTSVPS-PRSEATKXTGTGKP--TGTALPA 771  
 QY 217 STESSPTWPPCSEIPEKVKKPFDSQVYKFFISVFDPTTDHL 258  
 Db 772 ATQ--PVPVP--EPPAVAPAEPPVPDASGP-SPEPSHLL 808  
 RESULT 2  
 NCR2\_HUMAN STANDARD: PRT: 2517 AA.  
 ID NCR2\_HUMAN  
 AC Q9Y618; Q9Y500; Q1354; 000613; O15416;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Nuclear receptor co-repressor 2 (N-COR2) (silencing mediator of  
 DE retinoid acid and thyroid hormone receptor) (SMRT) (Thyroid-  
 DE retinoid-acid-receptor-associated co-repressor) (T3 receptor-  
 DE associating factor) (TRAC) (CTG26).  
 GN NCR2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RP [1]  
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
 RC TISSUE=pituitary;  
 RA MEDLINE=9917841; Pubmed=10077563;  
 RA Ordenlich P., Downs M., Xie W., Genin A., Splinter N.B., Evans R.M.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).  
 RN [12]  
 RN SEQUENCE FROM N.A. (ISOFORM SMRT).  
 RC TISSUE=cervical adenocarcinoma;  
 RA MEDLINE=99199215; Pubmed=10097068;  
 RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;  
 RT "SMRte, a silencing mediator for retinoid and thyroid hormone  
 RT receptors, a silencing isoform that is more related to the nuclear  
 RT receptor corepressor.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).  
 RN [13]  
 RN SEQUENCE OF 1023-2517 FROM N.A.  
 RC TISSUE=cervical adenocarcinoma;  
 RA MEDLINE=96008552; Pubmed=756127;  
 RA Chen J.D., Evans R.M.;  
 RT "A transcriptional co-repressor that interacts with nuclear hormone  
 RT receptors.";  
 RT Nature 377:454-457(1995).  
 RN [14]  
 RN SEQUENCE FROM N.A. (ISOFORM TRAC-1).  
 RC TISSUE=Fetal liver;  
 RA MEDLINE=96408715; Pubmed=8813722;  
 RA Sande S., Privalsky M.L.;  
 RT "Identification of TRACS (T3 receptor-associating cofactors), a family  
 RT of cofactors that associate with, and modulate the activity of,  
 RT nuclear hormone receptors.";  
 RT Mol. Endocrinol. 10:813-825(1996).

[5]  
 RP SEQUENCE OF 428-613 FROM N.A.  
 RC TISSUE-Brain cortex:  
 RX MEDLINE=97369492; PubMed=9225980;  
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,  
 Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
 RT "CDNAs with long CAG trinucleotide repeats from human brain.";  
 RL Hum. Genet. 100:114-122(1997).  
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TRITB.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: SMRT/TRAC-2 (SHOWN HERE) AND  
 CC TRAC-1; ARE PRODUCED BY ALTERNATIVE SPLICING. TRAC-1 CONTAINS ONLY  
 CC THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN  
 CC ANTI-REPRESSOR.  
 CC -1- TISSUE SPECIFICITY: UBQUITOUS. HIGH LEVELS OF EXPRESSION ARE  
 CC DETECTED IN LUNG, SPLEEN AND BRAIN.  
 CC -1- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.  
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
 CC FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CORN BOX.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, AF113003; AAD20946.1; -;  
 DR EMBL, AF125672; AAD22973.1; -;  
 DR EMBL, U37146; AAC50236.1; -;  
 DR EMBL, S83390; AAB50847.1; -;  
 DR EMBL, U80750; AAB91446.1; -;  
 DR MIM: 600848; -;  
 DR InterPro: IPR001005; Myb\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 2.  
 DR SMART: SM00395; SANT; 2.  
 DR PROSITE: PSS0090; MYB\_3; 1.  
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 KW Coiled coil; Alternative splicing.  
 FT DOMAIN 174 215 COILED COIL (POTENTIAL).  
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).  
 FT DNA\_BIND 429 474 SANT-A (POTENTIAL).  
 FT DNA\_BIND 613 657 MYB.  
 FT DOMAIN 522 561 COILED COIL (POTENTIAL).  
 FT DOMAIN 778 820 PRO-RICH.  
 FT DOMAIN 2139 2143 CORN BOX OF ID1.  
 FT DOMAIN 2342 2346 CORN BOX OF ID2.  
 FT DOMAIN 494 510 POLY-GLN.  
 FT DOMAIN 682 685 POLY-LYS.  
 FT DOMAIN 994 1002 POLY-PRO.  
 FT DOMAIN 1384 1389 POLY-PRO.  
 FT DOMAIN 1842 1846 POLY-GLY.  
 FT DOMAIN 2479 2482 POLY-PRO.

FT VARSPLOC 1 1702 MISSING (IN ISOFORM TRAC-1).  
 FT VARSPLOC 2353 2398 MISSING (IN ISOFORM TRAC-1).  
 FT CONFLICT 7 7 L -> P (IN REF. 2).  
 FT CONFLICT 295 295 K -> E (IN REF. 2).  
 FT CONFLICT 309 309 L -> W (IN REF. 2).  
 FT CONFLICT 352 352 MISSING (IN REF. 2).  
 FT CONFLICT 365 365 A -> P (IN REF. 4).  
 FT CONFLICT 612 613 SS -> EF (IN REF. 5).  
 FT CONFLICT 711 711 S -> T (IN REF. 2).  
 FT CONFLICT 724 740 MISSING (IN REF. 2).  
 FT CONFLICT 787 796 RTSRAPIEP -> PEDIPAPTES (IN REF. 2).  
 FT CONFLICT 804 804 G -> L (IN REF. 2).  
 FT CONFLICT 814 814 S -> F (IN REF. 2).  
 FT CONFLICT 817 817 A -> S (IN REF. 2).  
 FT CONFLICT 889 889 G -> R (IN REF. 2).  
 FT CONFLICT 1023 1030 SRSPAPPA -> MEAMAHF (IN REF. 3).  
 FT CONFLICT 1034 1034 A -> AKRPVEFA (IN REF. 2).  
 FT CONFLICT 1894 1894 K -> T (IN REF. 4).  
 FT CONFLICT 2494 2494 P -> A (IN REF. 4).  
 SQ SEQUENCE 2517 AA; 274031 MW; F58050C01761258C0 CRC64;  
 Query Match 8.0%; Score 119; DB 1; Length 2517;  
 Best local Similarity 20.2%; Pred. No. 1.1;  
 Matches 57; Conservative 33; Mismatches 134; Indels 58; Gaps 8;  
 QY 9 RDSFETTGSGSSVDLVGALPGLAPNLSSASVSASDSKKTIRPTTIKRSKES---- 64  
 DB 547 KEKDTDTSGEDNDEKEAASKGRKTANSQGRKRGRITSMANENANSEERITPOOSABELAS 606  
 QY 65 -----MSBOEHKFFLEALOLFDDRMKRTIEAFVQSTVQIHSKQYFLKYOKNCTR 116  
 DB 607 MELNSSWTEPEEMETAKGGLLEHGRNMSAIAKMGSTV-----SQCNEFFNFKKKNL 662  
 QY 117 EHVPPRPKRKASHPEYPOKASKNPVSOQVSTAPPTA-----ATOLDGCVYRAESSS 169  
 DB 663 DEILQOHKIMKEKRNARRKKKAPPAASEAPFPVEDEMEASGVSGNEBEKVEAE 722  
 QY 170 ILTKSGSCP-----TV-SSWVHTIPSIDASFVKD-----DGAPPG 206  
 DB 733 ALHAGNEVPKRGCEGSPATVNNSSDTEISIPSPHE-AKKDVGONGPKPPATGADGCPPG 781  
 QY 207 IET-----GNNCSGSGTSSPTWPCSEIPEKVP 237  
 DB 782 PPTPRKTSRAPIETPPASEATGATPPPAPEPSAPPPVP 823  
 RESULT 3  
 VNUA\_PRIVKA STANDARD; PRT: 1733 AA.  
 ID VNUA\_PRIVKA  
 AC P334B5;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Probable nuclear antigen.  
 OS Pseudorabies virus (strain Kaplan) (PRV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirus.  
 OX NCBI\_TaxID=33703;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021039; PubMed=2171211;  
 RA Vleck C., Kozmik Z., Paces V., Schlim S., Schyzer M.;  
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
 RT oriented open reading frame: characterization of their promoter and  
 RT enhancer regions.";  
 RL Virology 179:365-377(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC of EMBL: M34651: AAA47471.1; -

DR PIR: B45344; B45344.

FT DOMAIN 112 117 POLY-PHR.

FT DOMAIN 179 1733 GLY-RICH.

FT DOMAIN 192 196 POLY-SER.

FT DOMAIN 271 298 POLY-PRO.

FT DOMAIN 304 308 POLY-ARG.

FT DOMAIN 883 889 POLY-GLY.

FT DOMAIN 1398 1405 POLY-GLY.

SEQ SEQUENCE 1733 AA; 172166 MM; 0C8CB8BE475B5E2 CRC64;

Query Match 7.8%; Score 116.5; DB 1; Length 1733;  
 Best Local Similarity 26.8%; Pred. No. 1.1;  
 Matches 38; Conservative 15; Mismatches 56; Indels 33; Gaps 6;

QY 116 REHVPRPRPKASH-PYQKASKNV-PVSOQVSTAFPTAATQDSGY-----PRA 165  
 DB 93 RDHLPRRTRRDQHRRPPTTTTITKDPOHDPDLLPKTLQEDDPHLRPTRPDPSA 152  
 QY 166 ESSSILTKSSGSCPTVSSWVHTIPSDASFVEKKDQSGITGNMCSGTSSESPPT- 224  
 DB 153 KTHHHQDPGCGPSTSSHHHNDP-----PGGGPPSPPPRSTSSSHSGCPST 203  
 QY 225 -----WPCSEIPEKV 235  
 DB 204 RPPPPQRPRRWPPPS--PQKI 223

RESULT 4  
 BLM\_DROME STANDARD; PRT; 1487 AA.  
 AC 09VG18; 09Y062;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE Bloom's syndrome protein homolog (EC 3.6.1.-) (Dmblm) (Mutagen-  
 sensitive protein 309) (RecQ helicase homolog).  
 OS mus309 or blm or CG6920.  
 GN Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Phryganeola; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN 11  
 RP SEQUENCE FROM N.A., AND REPEATS.  
 RC STRAIN=Canton-S;  
 RX MEDLINE=99160561; PubMed=10049920;  
 RA Kusano K., Berres M.E., Engels W.R.;  
 RT "Evolution of the RECQ family of helicases: a Drosophila homolog,  
 RT Dmblm, is similar to the human Bloom syndrome gene.";  
 RL Genetics 151:1027-1039(1999).  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang J., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baellon R.M., Basu P., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Beus P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Broksstein P., Brotlier P.,  
 RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferltera S., Fleischmann W.,  
 RA Flosser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibbegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spiter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sviders R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: Participates in DNA replication and may participate in  
 CC repair. Exhibits a magnesium dependent ATP-dependent DNA-helicase  
 CC activity that unwinds single- and double-stranded DNA in a 3'-5'  
 CC direction (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.  
 CC -1- SIMILARITY: CONTRAINS 1 HRDC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entitles requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U92536; Amd1441.1; -  
 CC EMBL: AF003692; AAF54691.1; -  
 CC EMBL: F89n002306; mus309.  
 DR FlyBase; FBgn002306; DEAD.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002464; DEAH\_ATP\_helicase.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR002121; HRDC.  
 DR Pfam: PF00270; DEAD; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00570; HRDC; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 DR SMART; SM00341; HRDC; 1.  
 DR PROSITE: PS006590; DEAH ATP HELICASE; 1.  
 KW Hydrolyase; Helicase; ATP-binding; DNA replication;  
 KW Repeat; Nuclear protein.  
 FT DOMAIN 89 138  
 FT  
 FT REPEAT 89 112  
 FT REPEAT 115 138  
 FT DOMAIN 225 230  
 FT DOMAIN 464 467  
 FT DOMAIN 1283 1363  
 FT DOMAIN 1369 1372  
 FT DOMAIN 1416 1432  
 FT NP\_BIND 759 766  
 FT SITE 865 868  
 FT CONFLICT 98 98  
 FT CONFLICT 110 110  
 FT CONFLICT 126 126  
 FT CONFLICT 134 136  
 FT  
 FT 2 x 24AA REPEATS OF L-D-L-S-V-S-P-L-A-E-  
 FT L-[SP]-A-K-K-[YS]-[AD]-R-D-[SP]-P-P-K-  
 FT P.  
 FT 1.  
 FT 2.  
 FT POLY-PRO.  
 FT POLY-SER.  
 FT HRDC.  
 FT POLY-GLU.  
 FT NUCLEAR LOCALIZATION SIGNAL (BY  
 FT SIMILARITY).  
 FT ATP (BY SIMILARITY).  
 FT DEAH BOX.  
 FT K -> P (IN REF. 1).  
 FT K -> P (IN REF. 1).  
 FT L -> P (IN REF. 1).  
 FT SPK -> PPP (IN REF. 1).



```

DR Pfam: PF00023; ank: 4.
DR SMART: SM00248; ank: 2.
DR PROSITE: PSS0086; ANK_REPEAT: 1.
DR PROSITE: PSS0297; ANK_REPEAT_REGION: 1.
KM Coiled coil; Repeat; ANK repeat.
FT REPEAT 349 380 ANK 1.
FT REPEAT 384 413 ANK 2.
FT REPEAT 419 448 ANK 3 (POTENTIAL).
FT REPEAT 456 485 ANK 4.
FT DOMAIN 515 552 COILED COIL (POTENTIAL).
FT REPEAT 603 632 ANK 5 (POTENTIAL).
FT REPEAT 699 729 ANK 6 (POTENTIAL).
SQ SEQUENCE 919 AA; 100380 MW; SCE0022E0024EBA4 CRC64;

Query Match
Best Local Similarity 7.0%; Score 104.5; DB 1; Length 919;
Matches 69; Conservative 31; Mismatches 110; Indels 77; Gaps 12.

OY 22 DLYGMALPGIAPN---LSSASVYASA-----SEDAKIRKRPYITKSRMSQEHDKF 73
DB 582 DSVAKSKPVOEGIGYVLGSLSSARAPRKANDESDKILROLGKEISENVCTQKLSLE 641
OY 74 LEALOLFDDDWKTKIEFVSKRTVIOIRSHAKQYFLXKQKNTREHVPRRPKKASHPP 133
DB 642 FQOQAQSSNRSKKRIPIERKRELKLARLOLMQRSLESDTDSNNSQDPTTPVKRADPRP 701
OY 134 OKASKNVVPVSOQSTAFPTAAATQLDSGYPPRAESSILTKSGSSCPVSSMWHITP--- 190
DB 702 Q-----PIVESSES-----MDS-----AESLHLMIKR-----HTLASGG 730
OY 191 -----SIDASVEKEDKGPGPIETGNCCSGSTESS-----PTWPPCSEI 231
DB 731 RRPFPSTIKAS--KSLDGHSP-----SPTESESEPDLSEQYGSGISIPNPQ--SGD 777
OY 232 PERVKDEPFGSYVYKFGISVFPDPTTHLKLKLEWILQILKLCCTHEEP 278
DB 778 PQQSPDSTAAQKVATV--PKSAUKSPSSKRRTSQNLKLVTFEEP 821

RESULT 6
TBPL_YEAST STANDARD; PRT; 562 AA.
AC 002457;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TBPL protein (TTAGCG repeat-binding factor 1) (TBF alpha).
GN TBPL OR YPL128C OR IPI1C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=9932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LN24;
RX MEDLINE=93140769; PubMed=8423796;
RA Brigati C., Kurtz S., Balderes D., Vidali G., Shore D.M.;
RT "An essential yeast gene encoding a TTAGCG repeat-binding protein."
RL Mol. Cell. Biol. 13:1306-1314(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Hall J., Depaulo T., Ahmed A., Bussey H., Fortin N., Friese J.D.,
RA Storms R.K., Vo D.H., Wang Y., Mlnnett E.;
RA Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: BINDS TO TTAGCG REPEATS. ESSENTIAL FOR CELL GROWTH.
CC MAY PLAY A ROLE IN TELOMERE FUNCTION AND/OR STRUCTURE. MAY
CC REGULATE GENE EXPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: TO S. POMBE SPB1967.13.
CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X69394; CAA49191.1; -;  
 DR EMBL: U43703; AAB68230.1; -;  
 DR PIR: S28558; S28558;  
 DR PIR: A48079; A48079;  
 DR TRANSFAC: T01246; -;  
 DR SGD: S0006049; TRF1.  
 DR Pfam: PF00249; myb\_DNA-binding; 1.  
 DR SMART: SM00395; SANT; 1.  
 DR PROSITE: PS50090; MYB\_3; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 404 456 MYB.  
 FT CONFLICT 71 71 E -> R (IN REF. 1).  
 FT CONFLICT 371 373 DAA -> ERR (IN REF. 1).  
 SQ SEQUENCE 562 AA; 62823 MW; 2340F086468EC54F CRC64;

Query Match 6.9%; Score 102.5; DB 1; Length 562;

Best Local Similarity 20.7%; Pred. No. 2.9; Indels 55; Gaps 7;  
 Matches 45; Conservative 38; Mismatches 79;

QY 29 PGLAPNLSSASVSASASEDSAKK-----IRKPYRTKRSNSEQEHD 71  
 Db 363 PALQSIYDAVAASMSNSSGPHSHNSNSNNNSIGLRP---KAKRTMWSKEEE 418  
 QY 72 KFLFALDLPDRDMMKKEAFVG-----SKTVDIRSHAQYFLKVKNGTREVHP 121  
 Db 419 ALVEGLEKVEGPSMSKIDLYGPGGKITENLKNRYOLKOKARMKQLQYLSKG----K 472  
 QY 122 PRPKRASHPYPOKASKNVPSQOVSTAPPTATOLDGYYPRAESSTILTKSGSSCP-T 180  
 Db 473 PLP-----DYLKVTGNLEKIYAKKKFSQS-----PNSSTIMEONLSQHPSS 515  
 QY 181 VSSWVHHTIPSIDASFEVKDDGPGGIGETGNNGSSGS 217  
 Db 516 AASATEDTQTHQEDSHGONSDNMPNGLFGNSTSDNT 552

RESULT 7  
 EOMD\_XENLA STANDARD; PRT; 692 AA.  
 AC P79944;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Eomesodermin.  
 GN EOMES.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=gastrula;  
 RX MEDLINE=97133207; PubMed=8978604;  
 RA Ryan K., Garrett N., Mitchell A., Gurdon J.B.;  
 RT Eomesodermin, a key early gene in Xenopus mesoderm differentiation.\*;  
 RL Cell 87:989-1000(1996).  
 CC -1- FUNCTION: INVOLVED IN MESODERM DIFFERENTIATION. ACTIVATES WNT8,  
 CC BRACHYURY, CHD AND MIX.1 EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT OR JUST AFTER MIDBLASTULA  
 CC TRANSITION (STAGE 8). MAXIMALLY EXPRESSED AT STAGE 10 AS AN  
 CC EQUATORIAL MESODERM BAND, MORE PROMINENTLY ON THE DORSAL SIDE  
 CC AND AROUND THE INVAGINATING DORSAL LIP.  
 CC -1- INDUCTION: BY ACTIVIN.

CC -1- DOMAIN: CONTAINS 13 S-P-X-X REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 T-BOX DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U75996; AAC60061.1; -;  
 DR HSSP: P24781; 1XBR.  
 DR InterPro: IPR001699; T-box.  
 DR Pfam: PF00907; T-Box; 1.  
 DR PRINTS: PR00937; TBOX.  
 DR SMART: SM00425; TBOX; 1.  
 DR PROSITE: PS01283; TBOX\_1; 1.  
 DR PROSITE: PS01264; TBOX\_2; 1.  
 DR PROSITE: PS50252; TBOX\_3; 1.  
 KW Developmental protein; Transcription regulation; DNA-binding;  
 KW Nuclear protein; Repeat.  
 FT DNA\_BIND 263 443 T-BOX.  
 SQ SEQUENCE 692 AA; 75943 MW; 9D129A67F6357989 CRC64;

Query Match 6.8%; Score 101; DB 1; Length 692;

Best Local Similarity 21.4%; Pred. No. 4.7;  
 Matches 63; Conservative 37; Mismatches 95; Indels 100; Gaps 13;

QY 5 AETVRDSFETTSGGSSVDVGNALPGLAPNLSSASVSASASEDSAKKIRKPYRTKRSRS 64  
 Db 438 AKGFRNDYD-----SMYTASESDRLTPSPADSPSHOIVGTRRS 477  
 QY 65 MSEQEHKFLFALDLPDRDMMKKEAFVGSKTVYQLSHAQYFLKVKNGTREVHPPT- 123  
 Db 478 VQFFPDQOVNNL-----PRARYSGERTYPOANG-----LISQTEEVANPPQRM 525  
 QY 124 ---PKRKA-----SHP---YPOKA-----SK 138  
 Db 526 FVTPVQQAAMANKLDMGAVETDYSGLTYGIKSLPIQSHPMAYYDPAAFSMAGWGSR 585  
 QY 139 NVPSQOVSTAPPTATOLDGYYPRAESSTILTKSGSSCPVSSVHHTIPSIDASFE 198  
 Db 586 GSTYQRRKMTSLPWSRSSPSGPF---SEDLPLPKDKVKEEMSSWV-ETPISISL--- 636  
 QY 199 KDDGCGPGIETGN-----NCSGSTESSPTWPCSEIPEKVKDPFSQVYFFIG 247  
 Db 637 --DSNDSGYTTGACKRRRLSPSTSSNENSPPI--KCEDIGTE---DYKATRGIG 684

RESULT 8  
 FXPL\_MOUSE STANDARD; PRT; 705 AA.  
 AC P58462;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Forkhead box protein P1 (Forkhead-related transcription factor 1).  
 GN FOXP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A, B AND C).  
 RC STRAIN=C57BL/6; TISSUE=Lung;  
 RX MEDLINE=21347947; PubMed=11358962;  
 RA Shu W., Yang H., Zhang L., Lu M.M., Morrisey E.E.;  
 RT Characterization of a new subfamily of winged-helix/forkhead (Fox)  
 RT genes that are expressed in the lung and act as transcriptional  
 RT repressors.\*;  
 RL J. Biol. Chem. 276:27488-27497(2001).

```

CC -1- FUNCTION: Transcriptional repressor that play an important role in
CC the specification and differentiation of lung epithelium.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms: A (shown here), B and C; are
CC produced by alternative splicing. The isoform C is produced by the
CC use of alternative initiation codons in the same reading frame.
CC -1- TISSUE SPECIFICITY: Highest expression in the lung, brain, and
CC spleen. Lower expression in heart, skeletal muscle, kidney, small
CC intestine (isoform C not present) and liver.
CC -1- DEVELOPMENTAL STAGE: Expressed in developing lung, neural,
CC intestinal and cardiovascular tissues.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C2H2-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR D1 EMBL: AF339103; AAK69648.1; -
DR D1 EMBL: AF339104; AAK69649.1; -
DR D1 EMBL: AF339105; AAK69650.1; -
DR D1 MGD: MGI:1914004; Foxp1.
DR DR PROSITE: PS00657; FORK_HEAD_1; FALSE_NEG.
DR DR PROSITE: PS00658; FORK_HEAD_2; FALSE_NEG.
DR DR PROSITE: PS00339; FORK_HEAD_3; 1.
DR DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
DR DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Alternative splicing; Alternative initiation.
FT FT DNA_BIND 493 583 FORK-HEAD.
FT FT ZN_FING 334 359 C2H2-TYPE.
FT FT DOMAIN 55 60 POLY-GLN.
FT FT DOMAIN 71 107 POLY-GLN.
FT FT DOMAIN 161 164 POLY-GLN.
FT FT VARSPIC 1 250 MISSING (IN ISOFORM C).
FT FT VARSPIC 539 602 MISSING (IN ISOFORM B).
SQ SEQUENCE 705 AA; 78833 MW; 92962B82917CC79D CRC64;

Query Match 6.7%; Score 99.5; DB 1; Length 705;
Best local Similarity 22.8%; Pred. No. 6.2;
Matches 50; Conservative 26; Mismatches 92; Indels 51; Gaps 7;

QY 13 ETTSGG-SSVDLVGALDGLAPNLSSASVSASASEDSAKKIKRPYTTTRESWSBQEND 71
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
DB 271 ETTSSNHSSLDLTSTCVSSASPSSKSLIMNPHASTNGOLSVTP-----KRSLSHENP 325
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -

QY 72 -----KFLMALQLFDRDMKLEAFVSGSKVTYQIRSHQ 104
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
DB 326 HSHPLYGIGVCWPCGEAVCDPFAFLKHLNSENHALD-DRSTAQCRVMQVQOLEQLQA 384
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
QY 105 KYFLVQKNGTRHENVPPRPKRASHPYFQKASKNVPSQGVSTA---FPTAATQLDSG 160
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
DB 385 KDKERLQAMMTHTLHKSTEPK---AAPQPLNLVSSVTSLSKASEASPQSLPHPTPTTPAP 441
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -

QY 161 YYPRAESSILTKSGSSCPTVSSWVHRTIPSIDASFVEK 199
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -
DB 442 LTPVTQGPSVITTTSM-----HHVGPTRRRYS DK 470
   ||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - -

RESULT 9
ST20_YEAST STANDARD: PRT: 939 AA.
AC Q03497:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase STE20 (EC 2.7.1.-).
GN STE20 OR YHL007C.

```

OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N. A.  
RA MEDLINE=93099855; PubMed=1464311;  
RA Leberer E., Dignard D., Harcus D., Thomas D.Y., Whiteway M.;  
RT "The protein kinase homologue Ste20p is required to link the yeast  
RT pheromone response to protein kinase beta gamma subunits to downstream  
RT signalling components".  
RL EMO J. 11:4815-4824(1992).  
RN [2]  
RP SEQUENCE FROM N. A.  
RA MEDLINE=93133807; PubMed=8421676;  
RA Ramey S.W., Davis R.W.;  
RT "A dominant truncation allele identifies a gene, STE20, that encodes  
RT a putative protein kinase necessary for mating in Saccharomyces  
RT cerevisiae".  
RL Proc. Natl. Acad. Sci. U.S.A. 90:452-456(1993).  
RN [3]  
RP SEQUENCE FROM N. A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
RA Kucza T., Faveille A., Fulton L., Johnson S., Gaisel C., Kirsten J.,  
RA Dubez T., Hillier L., Jier M., Johnston L., Langston Y.,  
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,  
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RA Vaudin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII".  
RL Science 265:2077-2082(1994).  
CC -I- FUNCTION: REQUIRED TO LINK THE PHEROMONE RESPONSE G-PROTEIN BETA  
CC GAMMA SUBUNITS TO DOWNSTREAM SIGNALING COMPONENTS. IT IS THOUGHT  
CC THAT IT CAN PHOSPHORYLATE STE5. NEEDED FOR MATING IN HAPLOID  
CC CELLS, INDUCTION OF A MATING-SPECIFIC GENE FUS1, INDUCTION OF  
CC MATING-SPECIFIC MORPHOLOGIES, AND PHEROMONE-INDUCED PROLIFERATION  
CC ARREST. PHOSPHORYLATES STE11.  
CC -I- PATHWAY: RESPONSE TO PHEROMONE-INDUCED SIGNAL.  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -I- PTM: AUTOPHOSPHORYLATED ON SERINE RESIDUES.  
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC STE20 SUBFAMILY.  
CC -I- SIMILARITY: CONTRAINS 1 GBD DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on ways  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL, M94719; AAA35111.1; -;  
DR EMBL, L04655; AAA35038.1; -;  
DR EMBL, L04655; AAA35039.1; -;  
DR EMBL, U11581; AAB69747.1; -;  
DR PIR, S28394; S28394.  
DR PIR, S46821; S46821.  
DR HSDP, P11362; 1PGR.  
DR SGD, S0000999; STE20.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR000095; PAK\_box\_p21\_Rho\_binding.  
DR InterPro: IPR002290; Ser\_thr\_kinase.  
DR Pfam, PF00786; PBD; 1.  
DR SMART, SM00069; pkinase; 1.  
DR SMART, SM00285; PBD; 1.  
DR SMART, SM00220; S\_TKc; 1.  
DR PROSITE, PS50108; GBD; 1.  
DR PROSITE, PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE, PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE: PS00108: PROTEIN\_KINASE\_ST. 1.  
 KW Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 KM Phenomone response; Phosphorylation.  
 FT DOMAIN 337 395 GBD.  
 FT NP\_BIND 620 871 PROTEIN KINASE.  
 FT BINDING 649 634 ATP (BY SIMILARITY).  
 FT ACT\_SITE 739 739 ATP (BY SIMILARITY).  
 FT CONFLICT 19 19 N -> S (IN REF. 2).  
 FT CONFLICT 134 134 I -> M (IN REF. 2).  
 FT CONFLICT 271 271 P -> S (IN REF. 2).  
 SQ SEQUENCE 939 AA: 102362 MW: 69C1C12F5B87733C CRC64:

Query Match 6.6%; Score 98.5; DB 1; Length 939;  
 Best Local Similarity 20.8%; Pred. No. 10;  
 Matches 69; Conservative 44; Mismatches 123; Indels 95; Gaps 13;

QY 10 DSFETSGSSVDLVGM-----ALPGLAPNLSSASVSASDSAKKIRKP 55  
 DB 90 DPQPTKRVSSSVISGSSSSPHSNIDETKSLAVPNTNTMTDHSADN----- 142  
 QY 56 YTIKRSRESSEOEH---DKFLALQLFD-----RDMKKIEAFVGSKTIVQIRS--- 101  
 DB 143 ---TFSTINASESDHQFNDTLTKSLDSTETIENNATVAKHQCPVASSSTVNSKSSSTDI 199  
 QY 102 -----HAQYFLKVNKNGTRENHVPKPKKASHPYOKASK 138  
 DB 200 RRAATPVSTPVYISKPSMTTTPROINSASHSLSPKH--KQHKPKKPKSPKASKPVSVKK 257  
 QY 139 NVPSVQOQSTAFPTAATOLDGYYPRAESSILTKSGSCPTVSSWNHTIPSIDAFVE 198  
 DB 258 SFPKSNPLKNSNP-PKQOTEKSYT-----SSSKKRRKSGNSGTL-----RKMDVFTSEVQ 307  
 QY 199 -----KDDGPRPGIETGNCCSGSTESSPTTPPCSEIPEKVPKDFSOYKFTGSV-FD 251  
 DB 308 NIKRNSQDD-----KRASSSSNNNS-----SSSITALRISTPYNAKHHHGVGD 352  
 QY 252 PSTDHLKLLKLEWLIQILKLCCTHEPPHNL 282  
 DB 353 SKTGEYTGTPREWEKLLTSSGISKREQQNM 383

RESULT 10  
 ZDS2\_YEAST  
 ID ZDS2\_YEAST STANDARD; PRT; 942 AA.  
 AC P54786;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ZDS2 protein.  
 GN ZDS2 OR MCS1 OR YML109W OR YH8339.10.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96413276; PubMed=8816438;  
 RA Yu Y., Jiang Y.W., Wellinger R.J., Carlson K., Roberts J.M.,  
 RA Stillman D.J.;  
 RT "Mutations in the homologous ZDS1 and ZDS2 genes affect cell cycle  
 progression";  
 RL Mol. Cell. Biol. 16:5254-5263(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RL Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NOT KNOWN; MULTICOPY SUPPRESSOR OF A VARIETY OF DEFECTS.  
 CC SEEMS TO INTERACT WITH CDC42. MAY PLAY AN IMPORTANT ROLE IN CELL  
 CC CYCLE PROGRESSION.  
 CC -1- MISCELLANEOUS: ZDS MEANS 'ZILLION DIFFERENT SCREENS' AS BOTH ZDS1

CC AND ZDS2 HAVE BEEN FOUND BY A WIDE VARIETY OF GENETIC SCREENS.  
 CC -1- SIMILARITY: STRONG, TO YEAST ZDS1/NCI/CESI.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 CC EMBL: U32938: AAB37541.1; -;  
 DR EMBL: 249210; CAB89109.1; -;  
 DR SCD: S0004577; ZDS2.  
 FT CONFLICT 7 7 M -> V (IN REF. 1).  
 FT CONFLICT 23 23 T -> S (IN REF. 1).  
 FT CONFLICT 530 530 A -> P (IN REF. 1).  
 FT CONFLICT 546 546 A -> G (IN REF. 1).  
 FT CONFLICT 633 633 V -> A (IN REF. 1).  
 FT CONFLICT 668 668 N -> S (IN REF. 1).  
 FT CONFLICT 711 711 V -> VNC (IN REF. 1).  
 FT CONFLICT 723 723 MISSING (IN REF. 1).  
 FT CONFLICT 729 729 A -> E (IN REF. 1).  
 FT CONFLICT 833 833 S -> P (IN REF. 1).  
 FT CONFLICT 914 942 TGDIAFGDSALGAMDKNDSQGTILIPDI -> HMRYSLOW  
 (IN REF. 1).  
 SQ SEQUENCE 942 AA: 105495 MW: A1C9DD9A539E4291 CRC64;

Query Match 6.6%; Score 98.5; DB 1; Length 942;  
 Best Local Similarity 23.5%; Pred. No. 10;  
 Matches 73; Conservative 37; Mismatches 121; Indels 79; Gaps 15;

QY 6 EYR-----DSFETSGSSVDLVGMALPGLAPNLSSASVSASDSAKKI-----R 53  
 DB 549 QTYQQLDDEDECEVDNEKADFPVNLSPRAKSTRKASPRINRKRHSPIQIHSSEA 608  
 QY 54 KPTITPSRESSEOEHDK-----FLEALQLPDMKKITFAFGSKTVQIRSHAO----- 104  
 DB 609 KSVYITPSVSVSSSEQSPKPTAPAYVE-----KKVEL-----PTDQASTHKRNSLE 655  
 QY 105 -----KYFLKVNKNGT-----REHVPKPKKASH-----PYOKASKNVPSVQ 145  
 DB 656 KRLAKLFKRQHNQGTCKSDVKYIKKSV-KKELKKASHSLSKFRKSPKKKPEALEVERP 714  
 QY 146 VSTAFTPTAATOLD--SGYPRAESSILTKSGSCPTVSS-WVHHTIPSIDASFVKDDG 202  
 DB 715 SSPTKTTTTPEDIPYASVIEPRVSSNASTLLPDSHTSHSEPFVEFTISELDGDSFDISG 774  
 QY 203 GPFG--TETGNCCSGST-----ESSPTTPPCSEIPEK-----YKPDFSQ 241  
 DB 775 GDVNYDVEVHSSIRDTAGLEEDIGAEEDNTPSPAPQISTLPKRLTFEDVYKPDYSN 834  
 QY 242 V-YKFTGSVP 250  
 DB 835 APIKFTDSAF 844

RESULT 11  
 AT12\_HUMAN  
 ID AT12\_HUMAN STANDARD; PRT; 1593 AA.  
 AC P58397;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and  
 DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-  
 DE TS12).  
 GN ADAMTS12.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;



```

RT Redundant activation domains and interacts with TBP and TFIIB."
RL J. Biol. Chem. 273:23335-23343(1998).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LUNG-
CC SPECIFIC GENES.
CC -1- SUBUNIT: INTERACTS WITH THE TRANSCRIPTION FACTORS TBP AND TFIIB.
CC -1- TISSUE SPECIFICITY: Lung.
CC -1- DOMAIN: TWO ACTIVATION DOMAINS, AD1 AND AD2, C-TERMINAL OF (AND
CC DISTINCT FROM) THE FORKHEAD DOMAINS ARE NECESSARY FOR
CC TRANSCRIPTIONAL ACTIVATION.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF084939; AAD19875.1; -.
DR EMBL: AF084938; AAD19875.1; JOINED.
DR HSSP: 063245; 2HPH.
DR TRANSFAC: T02465; -.
DR MIM: 603350; -.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS00039; FORK_HEAD_3; 1.
DR DNAStrand: Nuclear protein; Transcription regulation; Activator.
FT DOMAIN 33 41 POLY-ALA.
FT DOMAIN 46 55 POLY-SER.
FT DOMAIN 58 67 POLY-SER.
FT DOMAIN 77 84 POLY-GLY.
FT DNAStrand 99 190 FORK-HEAD.
FT DOMAIN 262 272 POLY-HIS.
FT DOMAIN 301 306 POLY-GLY.
FT DOMAIN 312 315 POLY-SER.
FT DOMAIN 429 432 POLY-HIS.
SQ SEQUENCE 444 AA; 45993 MW; 32BDC5F373CFB147 CRC64;

Query Match 6.5%; Score 96.5; DB 1; Length 444;
Best Local Similarity 28.4%; Pred. No. 5.9; Mismatches 63; Indels 27; Gaps 6;
Matches 42; Conservative 16;

QY 115 TREHVPPRPKRKASHPYPOKASKNVPVSOVSTAFPTAATOLDGYPRAESSILTKS 174
DB 2 TTEGCPAPPLRRACSP-----VPGALQALMLSPPPAAAAAAAPPTSSSSSS 53
QY 175 GSSCPYVSSVWHHTIPSIDASFVEKDDGCPPIETG--NNCSCGSTRSSPTPPPCSET- 231
DB 54 ASCASSSSSSNSASAPs---AACSSAGGAGAGGAGGAGAKKASSGLRR--PEKPPSYIA 107
QY 232 -----PEKVPDFSOQYKFIGSVF 250
DB 108 LIYMAIOSPSRK-RLTLSEIYQFLQARF 134

RESULT 13
DLP3_RAT
ID DLP3_RAT STANDARD; PRT: 977 AA.
AC P97838;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Disks large-associated protein 3 (DAP-3) (SAP90/PSD-95-associated
DE protein 3) (SARAP3) (PSD-95/SAP90 binding protein 3).
GN DLAGP3 OR DAP3.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-9727335; PubMed-9115257;
RA Takachi M., Hata Y., Hirao K., Toyoda A., Irie M., Takai Y.;
RA "SARAPs. A family of PSD-95/SAP90-associated proteins localized at
RT postsynaptic density.";
RL J. Biol. Chem. 272:11943-11951(1997).
CC -1- FUNCTION: May play a role in the molecular organization of
CC synapses and neuronal cell signaling. Could be an adapter protein
CC linking ion channel to the subsynaptic cytoskeleton. May induce
CC enrichment of PSD-95/SAP90 at the plasma membrane.
CC -1- SUBUNIT: Interacts with Dlg1 and Dlg4/PSD-95 (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed in brain.
CC -1- SIMILARITY: BELONGS TO THE SARAP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67139; AAB48589.1; -.
DR KJ: Membrane.
FT DOMAIN 222 233 POLY-HIS.
SQ SEQUENCE 977 AA; 106019 MW; 1EFF13F5718C74FD CRC64;

Query Match 6.3%; Score 94.5; DB 1; Length 977;
Best Local Similarity 21.4%; Pred. No. 21; Mismatches 119; Indels 79; Gaps 12;
Matches 64; Conservative 37;

QY 7 TYRDSFEETSGGSSVDYLG--MALPGLAPNLSSASVSASA-----SEDSA 49
DB 293 SYRDLSEFFKRSRGSE---GRLACAGMSMLDQGVKRSAMHTMVGSGRDYPGAGCG 348
QY 50 KTRKRPYTTKRSKESSECEHDKFLDALQFDKMKKIAFYGSKTVIOIRSHAKYFLK 109
DB 349 KGLIGPETHAKART-----YHYLDVPPDDMGVYPT-GGKDELPCRRMRSGSYIK 397
QY 110 V---QKNGTREHVPPRPKRKASHPYPOKASKNVPVSO----- 144
DB 398 AMODESGSDGSPKTSPPALARR-FASRRSSVDGTARINCCVPRRIHPRSSIPGYSNL 456
QY 145 ---QVSTAFPTAATQDSCGYYPRAESSILTKSGSCPTVSSVWHHTIPSIDAFYEKDD 201
DB 457 TTCQLSEEFNQOLEAVCGSVFGELESQAVDALDLPCCFMRSS--HSYLAIAIGCSQDD 514
QY 202 GGP---PGIETGNCCSGSTESSPTPPPCSEIPEKVPDFDSQYKFIGSVDFDSTTD 256
DB 515 CLPILAAPASVSGRPGSSFNFRKAPPPIPPGSAAPPRI-----SITAGSSTD 561

RESULT 14
PAX6_BRARE
ID PAX6_BRARE STANDARD; PRT: 437 AA.
AC P26630;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Paired box protein PAX1(ZF-A) (PAX-6).
GN PAX6A OR PAXZF-A OR PAX1(ZF-A).
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.

```

```

OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92037521; PubMed=1718739;
RA Krauss S., Johansen T., Korzh V., Moens U., Ericson J.U., Fjose A.;
RT "zebrafish pax1(zf.a); a paired box-containing gene expressed in the
RL neural tube."
RN EMBL J. 10:3609-3619(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=91375540; PubMed=1660220;
RA Krauss S., Johansen T., Korzh V., Fjose A.;
RT "Expression pattern of zebrafish pax genes suggests a role in early
RL brain regionalization."
RN Nature 353:267-270(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315909; PubMed=1352238;
RA Puschel A.W.P., Gruss P., Westerfield M.;
RT "Sequence and expression pattern of pax-6 are highly conserved
RL between zebrafish and mice."
RL Development 114:643-651(1992).
CC -1- FUNCTION: TRANSCRIPTION FACTOR EXPRESSED IN SPATIALLY RESTRICTED
CC REGIONS OF THE NEURAL TUBE DURING EMBRYONIC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM AND A SHORT FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SPATIALLY RESTRICTED REGIONS OF THE NEURAL
CC TUBE.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X61389; CAA43661.1; -
DR EMBL: X63183; CAA44867.1; -
DR PIR: S18558; S18558.
DR HSSP: P26367; 6PAX.
DR TRANSFAC: T00682; -
DR ZFIN: ZDB-GENE-990415-200; pax6a.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR001523; Paired_box.
DR Pfam: PF00046; homeobox; 1.
DR Pfam: PF00292; PAX; 1.
DR PRINTS: PRO0027; PAIREDBOX.
DR SMART: SM00389; HOX; 1.
DR SMART: SM00351; PAX; 1.
DR PROSITE: PS00027; HOMEBOX.1; 1.
DR PROSITE: PS00071; HOMEBOX.2; 1.
DR PROSITE: PS00034; PAIRED_BOX; 1.
KW Transcription regulation; Paired box; Homeobox; DNA-binding;
KW Nuclear protein; Developmental protein; Alternative splicing.
FT DOMAIN 23 147 PAIRED BOX.
FT DNA_BIND 229 288 HOMEBOX.
FT DOMAIN 298 437 SER/THR/PRO-RICH.
FT VARSPIC 66 66
SQ SEQUENCE 437 AA; 46425 MW; 01D0999EFC22837 CRC64;

```

```

OY 9 RDSFTTSGGSSVDLVGMALDGLAPLNSASVSASASSEDNAKKTRKPYTTKRSRNSQ 68
Db 193 QDGCGDSKGGGE-----NTNNTSSNGEDSDETQMRLQLRKLRNRTSTFQE 239

Query Match 6.3%; Score 94; DB 1; Length 437;
Best Local Similarity 19.9%; Pred. No. 8.7;
Matches 55; Conservative 54; Mismatches 91; Indels 76; Gaps 14;

OY 69 EHDKFLFALQLRDMKKI---EAFVGSKTVIOI---RSHAQKYLKYKNGCTREHPP 122
Db 240 Q-----IEAL-----EKEPERHYPDVFAERERLAIDLPERLIQVWFSNRKARRRE-KLR 291
OY 123 RPKRKASHPPKPKASKNVVSOOVSTAF-----PTAATQLDSGYPRAESSSILTKSGS 176
Db 292 NGRQASN-----SSSHIRISSFSFSTVYQPIPOPTTPPSFTSGSM-LGRSDTALTNTYS 345
OY 177 SCPYVSSW-----VHHTIPSIDASEYKDDGPPGTEGNN-----CS 214
Db 346 ALPPMPFSFTMANNLPPQPSQTSYSCLPTSPSVNGR--SYDTYPPPHQAHMNSQMAA 403
OY 215 SGSTES-----SPPTPCSEIPEKVKPDPSQOY 243
Db 404 SGTSTGTLISPGVSVPVGVGSE-----PDMSQY 433

RESULT 15
OTX3_BRARE
ID OTX3_BRARE STANDARD; PRT; 338 AA.
AC 090267;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Homeobox protein OTX3 (ZOTX3).
GN OTX3.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON AB;
RX MEDLINE=95206106; PubMed=7898305;
RA Mori H., Miyazaki Y., Morita T., Nitta H., Mishina M.;
RT "different spatio-temporal expressions of three otx homeoprotein
RL transcripts during zebrafish embryogenesis."
RL Brain Res. Mol. Brain Res. 27:221-231(1994).
CC -1- FUNCTION: MAY PLAY A ROLE IN VERY EARLY EMBRYOGENESIS,
CC GASTRULATION, AND THE DEVELOPMENT AND SUBDIVISION OF THE
CC Diencephalon AND THE MIDBRAIN. MAY PLAY A ROLE IN THE ORGANIZER
CC FUNCTION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: DETECTED ALONG THE MARGINAL ZONE OF SYMMETRIC
CC EMBRYOS AT 5 HRS OF DEVELOPMENT AND FOUND AT THE SHIELD, A PRIMARY
CC MORPHOLOGICAL ASYMMETRY AT 6 HRS OF DEVELOPMENT. FOUND IN TWO
CC STRIPES AT THE POSTERIOR SIDE OF THE EYE rudiments AT 12 HRS OF
CC DEVELOPMENT. DISTRIBUTED IN THE Diencephalon, MIDBRAIN AND THE
CC EPIDERMIS AT 18 HRS OF DEVELOPMENT, AND FOUND IN THE Diencephalon
CC AND THE MIDBRAIN AT 24 HRS OF DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC "BICOID" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D26174; BAA05160.1; -
DR HSSP: P06601; 1FJL.
DR ZFIN: ZDB-GENE-980526-27; otx3.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR003025; Otx_TF.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PRO0024; HOMEBOX.
DR PRINTS: PRO1255; OTXHOMEBOX.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX.1; 1.

```

DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 FT DNA\_BIND 38 97 HOMEBOX.  
 FT DOMAIN 95 99 POLY-GLN.  
 FT DOMAIN 134 144 POLY-SER.  
 FT DOMAIN 158 162 POLY-SER.  
 SQ SEQUENCE 338 AA; 36145 MW; ACFR6C2A43A69014 CRC64;

Query Match 6.2%; Score 93; DB 1; Length 338;  
 Best Local Similarity 19.9%; Pred. No. 7.6;

Matches 48; Conservative 40; Mismatches 103; Indels 50; Gaps 9;

QY 25 GMLPLGLAPNLSASVSASASEDSAKIRKPYTTKSGRESWSEQHDKFLDALQLFDRDW 84  
 Db 14 GLGLGATMDLHPVGVPTNPKQRRTTFTRTQ-----LDILESLFAKT 60  
 QY 85 KKIFAFYGSKTVIOI---RSHAQKF-----LKVQKNGTREHVPPRPKRKASHYP 133  
 Db 61 RYPIEFMREEVALKINLPESHVQVWFKNRRKACRQOQOQSSSTNSKIRPAKKPPSPRES 120  
 QY 134 QKASKNPVVSQOVSTAFPTAATQLDGVPYPAESS-SILTKSGSSCPTVSSWVHTIPSI 192  
 Db 121 GSSESGHFTTPRAVSS--SSSSSSSSGNSPALSGLVGLISSSSGCTVPSPIMSPAFYSPV 178  
 QY 193 DASFEKDDGGPPGI-----ETGNNC-----SSGSTSSPPTWPPCSEIPEKYKPDFSQY 242  
 Db 179 PA-----PPLPDISPASASCMQRAMSSGGTGTGVPSYP---MPYNOAFSYAQG 225  
 QY 243 Y 243  
 Db 226 Y 226

Search completed: August 12, 2002, 10:12:32  
 Job time: 225 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:08:27 ; Search time 26.58 Seconds  
(without alignments)  
1902.077 Million cell updates/sec

Title: US-09-640-211a-1076  
Perfect score: 1492  
Sequence: 1 MPMLAEYRDSFEFTSGSS.....EMIQILIKLCTHEEPFHL 282

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 539846 seqs, 179280859 residues

Total number of hits satisfying chosen parameters: 539846

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/prodata/1/paa/PCOT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/prodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701	47.0	336	6	US-10-155-881-21296
2	700	46.9	318	6	US-10-155-881-9544
3	700	46.3	325	6	US-10-155-881-20394
4	675.5	45.3	277	6	US-10-155-881-32157
5	664	44.5	304	6	US-10-155-881-32154
6	664	44.5	321	6	US-10-155-881-21324
7	628	42.1	313	6	US-10-155-881-28543
8	621	41.6	260	6	US-10-155-881-32135
9	618	41.4	333	6	US-10-155-881-28437
10	609.5	40.9	270	6	US-10-155-881-21319
11	585.5	39.2	231	6	US-10-155-881-10664
12	581	38.9	329	6	US-10-155-881-32140
13	578.5	38.8	187	6	US-10-155-881-9536
14	574.5	38.5	293	6	US-10-155-881-32148
15	550.5	36.9	214	6	US-10-138-145-1187
16	520.5	34.9	271	6	US-10-155-881-32151
17	520.5	34.9	359	6	US-10-155-881-9550
18	520	34.9	214	6	US-10-155-881-25350
19	519.5	34.8	229	6	US-10-155-881-32155
20	492	33.0	167	6	US-10-155-881-20389
21	489	32.8	181	6	US-10-155-881-21297
22	488.5	32.7	178	6	US-10-155-881-21320
23	481.5	32.3	174	6	US-10-155-881-21468
24	462	31.0	216	6	US-10-155-881-32691
25	460	30.8	207	6	US-10-155-881-10666
26	370.5	24.8	630	6	US-10-155-881-28422

27	367.5	24.6	267	6	US-10-155-881-18237	Sequence 18237, A
28	363	24.3	237	6	US-10-155-881-28014	Sequence 28014, A
29	363	24.3	260	6	US-10-155-881-28015	Sequence 28015, A
30	341	22.9	377	6	US-10-155-881-32144	Sequence 32144, A
31	339	22.7	169	6	US-10-155-881-28546	Sequence 28546, A
32	332	22.3	664	6	US-09-573-655B-1809	Sequence 1809, A
33	332	22.3	725	6	US-10-155-881-10033	Sequence 10033, A
34	331	22.2	438	6	US-10-155-881-10698	Sequence 10698, A
35	324	21.7	382	6	US-10-155-881-21365	Sequence 21365, A
36	324	21.7	387	5	US-09-935-625-5230	Sequence 5230, Ap
37	324	21.7	387	5	US-09-935-625-7742	Sequence 7742, Ap
38	324	21.7	481	6	US-10-155-881-28417	Sequence 28417, A
39	323.5	21.7	287	6	US-10-155-881-21366	Sequence 21366, A
40	323	21.6	557	6	US-10-155-881-25695	Sequence 25695, A
41	321	21.5	723	6	US-10-155-881-10030	Sequence 10030, A
42	320.5	21.5	745	6	US-10-155-881-26940	Sequence 26940, A
43	320	21.4	432	6	US-10-155-881-10671	Sequence 10671, A
44	312	20.9	287	5	US-09-935-625-1011	Sequence 1011, A
45	312	20.9	287	5	US-09-935-625-8409	Sequence 8409, Ap

## ALIGNMENTS

```
RESULT 1
US-10-155-881-21296
; Sequence 21296, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyeva, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 21296
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Glycine max
; US-10-155-881-21296

Query Match 47.0%; Score 701; DB 6; Length 336;
Best Local Similarity 58.6%; Pred. No. 5, 2e-46;
Matches 153; Conservative 24; Mismatches 58; Indels 26; Gaps 7;

OY 22 DLYGMLPG---LAPNLSASVSASASDSAKIRKPYRTTSRESSEODHDKFILEAO 78
Db 20 DPSNMPLPGVNNLPPPPPPAPAAVEDPNKIRKPYRTTSRESWTDGHDHDKFILEAO 79
OY 79 LFPDRMKKIEAFVGSXTVIOIRSHAQKYLKYNKCTREHVPPRRKRASHPYPOKASK 138
Db 80 LFPDRMKKIEAFVGSXTVIOIRSHAQKYLKYNKCTSEIVPPRRKRRAAHYPQKAKR 139
OY 139 NPEVSO-----QVSTAPPTAATOLDGSGYYPRAESSILTKSGSCPTVSSGWHHTPTST- 192
Db 140 TPTVSOVGMLPGSSSAF-----IEPAYIYSPSSSVLGTPTVYNNMP-LSSMWNNTTPQPG 192
OY 193 DASFVEKDDCGPPGCIENG---NNCSSGSTPSSPPTPSPCEI-----PEYKKPQFSOV 242
Db 193 NPDVTRDDMGILGACGAADALNCCYSSSNBSPPTWPRSKRINOGDQGRKIVKMPDAFV 252
OY 243 YKFIGSVFDPSTIDHLKIKE 263
Db 253 YSFIGSVFDPNSTNHLKIQ 273

RESULT 2
US-10-155-881-9544
```

```
: Sequence 9544, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 9544
: LENGTH: 318
: TYPE: PRT
: ORGANISM: Zea mays
US-10-155-881-9544
```

```
Query Match          46.9%; Score 700; DB 6; Length 318;
Best Local Similarity 59.0%; Pred. No. 5.8e-46;
Matches 148; Conservative 26; Mismatches 45; Indels 32; Gaps 7;
```

```
OY 21 VDLVGMALPGIAPNLSSASVSASSEDGAKIRKPYITTKSRRESWDEQEHKFLFALQLF 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 34 LSNVSASAP--PPPOAGSDAAGSEESKRVKRPYITTKSRRESWDEQEHKFLFALQLF 91
OY 81 DRDMKRIAPVGSKTIVDIRSHAOKYFLKYOXNGTREHVPKRRKASHPYPOKASKNV 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 92 DRDMKRIEAFVGSKTIVDIRSHAOKYFLKYOXNGTSEHVPPRRKRAAHYPPOKASKNE 151
OY 141 PVSQOVSTAFPTAATQLDSGYTPRAESSILTKSGSSCPTVSSVHHHTIPSIDASFVEKD 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 152 P-----NYGLKTDSSSIHRNSGMNV-SVSSWPHRSIQAVASSMWKD 192
OY 201 DG-GPGCIETGNNCSTESSPTTPW-----PCSEIPE-KVMPDPSQYKFTGVSFDP 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 193 LGAGTPEG--PNNFCSSSTEGPPTWPGETNGPFIQLPDLMPDRAGVYSFLGSVFPD 249
OY 253 STTDHLKRLKE 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 250 STSDHLQRLKE 260
```

```
RESULT 3
US-10-155-881-20394
: Sequence 20394, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 20394
: LENGTH: 325
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-20394
```

```
Query Match          46.9%; Score 700; DB 6; Length 325;
Best Local Similarity 57.9%; Pred. No. 6e-46;
Matches 143; Conservative 33; Mismatches 49; Indels 20; Gaps 5;
OY 25 GMAIPGLAPNLSSASVSASSEDGAKIRKPYITTKSRRESWDEQEHKFLFALQLFDRD 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 43 GMLPGMPPPAVSAAATAANSPEDDAKKIRKPYITTKSRRENWTEPEHDKFLFALQLFDRD 102
OY 84 WKRIEAFVGSKTIVDIRSHAOKYFLKYOXNGTREHVPKRRKASHPYPOKASKNVPS 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 WKRIEAFVGSKTIVDIRSHAOKYFLKYOXNGTSEHLPKRRKRAAHYPPOKASKNAPVL 162
OY 144 QOVSTAFPTAATQLDSGYTPRAESSILTKSGSSCPTVSSVHHHTIPSIDASFVEKDDG 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 163 SOVSAGFOSSSALLIEPGYILKNDAPML-KTPIIMNTVASSMSNNTLOTANLS----- 213
OY 204 PPGCIETGNNCSTESSPTTPWPCSE-----IPEKVPDPSQYKFTGVSFDPSTTD 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 214 -PYTKVNNPCSSG--ESTPWPVPGESNGGKNKIHPLRVLPDPTQYVGFISVFDNATE 270
OY 257 HLKRLKE 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 HLQRLKE 277
```

```
RESULT 4
US-10-155-881-32157
: Sequence 32157, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 32157
: LENGTH: 277
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-32157
```

```
Query Match          45.3%; Score 675.5; DB 6; Length 277;
Best Local Similarity 58.1%; Pred. No. 3.7e-44;
Matches 136; Conservative 34; Mismatches 45; Indels 19; Gaps 4;
```

```
OY 37 SASVSASASEDSAKIRKPYITTKSRRESWDEQEHKFLFALQLFDRDMKRIEAFVGSKTIV 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 8 SAATAANSPEDDAKKIRKPYITTKSRRENWTEPEHDKFLFALQLFDRDMKRIEAFVGSKSV 67
OY 97 IOIRSHAOKYFLKYOXNGTREHVPKRRKASHPYPOKASKNVPSQOVSTAFPTAATQ 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 IOIRSHAOKYFLKYOXNGTSEHLPKRRKRAAHYPPOKASKNAPVL SOVSAGFOSSSAL 127
OY 157 LDGSGYTPRAESSILTKSGSSCPTVSSVHHHTIPSIDASFVEKDDGCPGCIETGNNC 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 128 LEPGYILKNDAPML-KTPIIMNTVASSMSNNTLOTANLS-----PYTKVNNPCSSG 177
OY 217 STSSPTTPWPCSE-----IPEKVPDPSQYKFTGVSFDPSTTDHLKRLKE 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 -ESTPWPVPGESNGGKNKIHPLRVLPDPTQYVGFISVFDNATEHLQRLKE 229
```

```
RESULT 5
US-10-155-881-32154
: Sequence 32154, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 32154
: LENGTH: 277
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-32154
```

```

1 FILE REFERENCE: 38-21(15300)J
2 CURRENT APPLICATION NUMBER: US/10/155,881
3 CURRENT FILING DATE: 2002-05-22
4 NUMBER OF SEQ ID NOS: 37555
5 SEQ ID NO 32154
6
7 LENGTH: 304
8
9 TYPE: PRT
10 ORGANISM: Glycine max
11
12 US-10-155-881 -32154

```

Query Match	44.5%	Score 664	DB 6	Length 304
Best Local Similarity	51.3%	Pred. NO. 3.2e-43		
Matches 138	Conservative 42	Mismatches 35	Indels 34	Gaps 5

QY	21	VLDGMLPLGLAPLMLSSASVSASASESBAKKIRKRPYTITSRSMSQOEHDKLFLEALOLF	80
Db	14	LDPGSMTLPGLPPLAAATATAATSEEDPAKTRKPYITTSRSBSWEPEHDKLFLEALOLF	73
QY	81	DROMKKEAFEGSKTVIOLIRSHAQKYELKVOKNCTREHVPPRRKRASHPYOLKASKNV	140
Db	74	DROMKKEAFEGSKTVIOLIRSHAQKYELKVOKNCTREHVPRLPRKRKAARPYOKASKTA	133
QY	141	PVSOOVSTAFFTAATOLDGTYTPRAESSILTKSGSSCPVTSSWVHTTDSIDASFYEKD	200
Db	134	PVLQSVSGSFQSSALLPEGYILKHDSNA-MPKRPIINTAVSSWSNLSLOKTSTVLHGOK	192
QY	201	DGGPGIEGTGNCCSGSTESSEPWPWCSEL-----PEAKRPDFOVYKRTIGS	248
Db	193	-----QKVNNCCSSSKS-----PRAOLDGESNGORNNSHPLRALPFDAEYSTIGS	238

```

RESULT      6
US-10-155-881-21324
: Sequence 21324, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotsen, Stanton B.
: APPLICANT: Kovacic, David K.
: APPLICANT: Liu, Jingtong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIlinch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37555
: SEQ ID NO 21524
:
: LENGTH: 321
:
: TYPE: PRT
: ORGANISM: Glycine max
: US-10-155-881-21324

```

	Query Match	44.5%	Score 664	DB 63	Length 321
	Best Local Similarity	51.3%	Pred. No. 3.4e-43		
	Matches 138	Conservative 42	Mismatches 53	Indels 34	Gaps 5
Qy	21	VDLVGMALPGLGLPNLSASVSASASASASASAKTRRKYTTIKRSRWSGEQDFLEALOLF	80		
Db	31	LDPSGMSLPGILPFAAATATADSFEDPAKTRRKYTTIKRSRWSRTERPDKFLFALOLF	90		
Qy	81	DDDMKKIEAFVSGSKTVTIQIRSHAQKYLELVOKNGTRENHPVPPRPKKASHPPYQKASKNV	140		
Db	91	DDDMKKIEAFVSGSKTVTIQIRSHAQKYLELVOKNGTRENHLPVPPRPKKAAHPYQKASKTA	150		
Qy	141	PVSGQVSTAFPPAAATQDLSGYTPRAESSILTKSGSCCTVSSVNHHTLPISDASVEVD	200		
Db	151	PVLSGVSGFSSSALLGECYTIKIDHSSA-MKRTPLINTAVSSWNSSNLOKTTVSYLHGCK	209		

```

Oy 201 DGGPGIETGNCSSGSESPFWPCSEI-----PEKYPPOSYVYFIS 248
      ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 210 -----QKVNCCSSSS-----FRAOLVGSNGQRNNHPRLVDPFAEYSPIS 255
      | | | | | | | | | | | | | | | | | | | | | | | |
Oy 249 VFDPSTDLKKL-----EMLIQLIK 270
      ||| | | | | | | | | | | | | | | | | | | | | |
Db 256 VFDPVYTGHVOKLRMPDIDVEYLLMR 284
      ||| | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-10-155-881-28543
: Sequence 28543, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dolson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ. ID NOS: 37595
: SEQ ID NO 28543
: LENGTH: 313
: TYPE: PRT
: ORGANISM: Oryza sativa
US-10-155-881-28543

```

Query Match	42.1%	Score 628:	DB 6:	Length 313:
Best Local Similarity	48.9%	Pred. No. 1,9e+40:		
Matches 139:	Conservative 19:	Mismatches 40:	Indels 86:	Gaps 6:

  

QY	33	PULSASASASASDESAKKIRKPYTTTKRESMSDEHDKFLTEALQUL-----	79
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	8	PDAAAAAAGSAGEASKKVRKRPYTTTKRESMTDEHDKFLTEALDINLEYMLPTYMS	67
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
QY	80	-----PRDMKRIEAFGSKTVIOIRSHAQKRYLK	109
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	68	PYLMIANNVFMRTKNTAFVAVEFKLYLSPFRDMKKLEAFGSKTVIOIRSHAQKRYLK	127
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
QY	110	VOKNGTREVPPRRPKRKAASHPYPOKASKNVPSQOVSATFPAATOLDGTYIPRAESS	169
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	128	VOKNGTSEIVPPRRPKRKAASHPYPOKASKNP-----GYTKADSSS	169
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
QY	170	ILTKSGSSC-----PTVSSWNHTIPSDASFEVCKDDGGPGLGTGNKSSGSTE	219
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	170	MLRELRLHEHCVMDOGFPTNCSLIH-----GERFVVRDLASA--GAAPANNFGCSNTE	221
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
QY	220	SSPPTWPCSELPREKVPKPFPSQYKFGVFPDSTTDHLKLKE	263
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	222	GPAPAMQP-----VMPDFAQYVSFLGVSFPPSTSGHLQKLKE	258
		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :	

```

RESULT      8
US-10-155-881-32135
; Sequence 32135, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovacic, David K.
; APPLICANT: Liu, Jindong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 32135

```

```

? LENGTH: 260
? TYPE: PRT
? ORGANISM: Glycine max
? FEATURE:
? NAME/KEY: (1)..(260)
? LOCATION: unsure
? OTHER INFORMATION: unsure at all Xaa locations
US-10-158-881-32135

```

Query Match	41.68	Score 621	DB 6	length 260
Best Local Similarity	55.68	Pred. NO. 5.2e-40		
Matches 125, Conservative	29	Mismatches 61	Indels 10	Gaps 3

```

QY 46 EDSAKKIRKPYTITKSSSEWEOEHKFLLEALOLFPRODKKIEAFGSGTVOIINSHAK 105
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2 EDOKKKYRKRYTITKSSSEWNTDOEHKFLLEALHFLPRODKKIEAFGSGTVOIINSHAK 61

QY 106 YFLKQKNGTRENHPRPBRKASHPYPOKASKNVPVSOOVSTAEPTATOLDSCGYBRA 165
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 62 YFMKQKNGTSEHNPPRPBRKKAHPRPKASKNALTISOVARPLQSSSALSSSHIYRP 121

QY 166 ESSSILTKSSGSCPYSSWVHHITPISLIDASEVEKDDGCPPIETGNCCSGSTESSPPTW 225
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 122 DSSSVVTRTPVROCGX-PSMVGYNTPPLVSLJRPRTKDD--HVLMSQOINPFSSSENEHTPRGM 178

QY 226 PRCEI-----PEKYKPRDSOVYKFKISVYCPSTTIDLLKLKE 263
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 179 PISKQTDGOGDGOKPTIYMPDFAGVYSFICITVEPNNIINLQRIKQ 223

```

RESULT 9  
US-10-155-881-28437  
: Sequence 28437, Application US/10155881  
: GENERAL INFORMATION:  
: APPLICANT: Dotson, Stanton B.  
: APPLICANT: Kovacic, David K.  
: APPLICANT: Liu, Jingdong  
: APPLICANT: Lutfiyya, Linda L.  
: APPLICANT: McIninch, James  
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
: FILE REFERENCE: 38-21(15300)J  
: CURRENT APPLICATION NUMBER: US/10/155,881  
: CURRENT FILING DATE: 2002-05-22  
: NUMBER OF SEQ. ID NOS.: 37595  
: SEQ. ID NO 28437  
: LENGTH: 333  
: TYPE: PRT  
: ORGANISM: *Oryza sativa*  
US-10-155-881-28437

Query Match 41.4%; Score 618; DB 6; Length 333;  
Best Local Similarity 50.0%; Pred. No. 1.2e-39;  
Matches 136; Conservative 37; Mismatches 63; Indels 36; Gaps 10

Qy	17	GGSSVDIYGMALPLAANLSSASVSASASDSKAKRKYTTTKSRSESDREHKKFLEA	76
Dd	33	GGRKEKEQVYAAP--LOPPMAVPAPAAAVGEE-ARKRKRYTTTTSKRSTREPEHHKFLEA	90
Qy	77	LQLFDROMKKIIEAEVGSRTVIQRSHAQKLYLKVNOKNGTRHEHVPPRPKRKAHPYPOKA	136
Dd	91	LQLFDROMKRIEAAVGVSGSKTVIQIRSHAQKFLYLKVGKNGTGEHLPPRRPKRAHPYPCKA	150
Qy	137	SKNPVPSGOQVSTAFTPATOL-----DSGYIPRAESSILITKGSSCPHYSSVMHTTISI	192
Dd	151	SKNVS-----PAAISQPRLPGEGCCVMSMDTSPVLRNTASA-VVPSMDSNTIAOPL	200
Qy	193	DASFVEKDGDGPPIELGNNGSCGSTRESSPMPCESEI-----PEKVARDPEQVKKF	245
Dd	201	SASTTO-----GTGAIVAT--NKCSS-STIESBSTTWPISLEAYEOENMLRPLAMDEVAQVYSF	254

```
QY 246 IGSVFDPSTTDHLKLLK-----EWIQLILK 270
      :||:| | | : || : | | : | : :
Db 255 IGSIFDPDTSCHLQTLKAMPIDIVETVLLMR 286
```

```

RESULT 10
US-10-155-881-21319
: Sequence 21319, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovacic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ. ID NOS: 37595
: SEQ. ID NO 21319
: LENGTH: 270
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-21319

```

Query Match	40.9%	Score 609.5	DB 6	Length 270
Best Local Similarly	55.0%	Pred. No. 4.2e-39		
Matches 132, Conservative	27	Mismatches 58	Indels 23	Gaps 5

[illegible]

```

RESULT 11
US-10-155-881-10664
: Sequence 10664, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovacic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ. ID NOS: 37595
: SEQ. ID NO 10664
: LENGTH: 231
: TYPE: PRT
: ORGANISM: Zea mays
US-10-155-881-10664

```

Query Match	39.28;	Score 585.5;	DB 6;	Length 231;
Best Local Similarity	53.38;	Pred. No. 2.4e37;		
Matches 130; Conservative	20;	Mismatches 57;	Indels 37;	Gaps 7;

[illegible]

```

RESULT 14
US-10-155-881-32148
; Sequence 32148; Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovacic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIntoch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 32148
; LENGTH: 293
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
US-10-155-881-32148

```

Query Match	38.5%	Score 574.5	DB 6	Length 293:
Best Local Similarity	47.0%	Pred. No. 2.3e-36:		
Matches 133:	Conservative 17:	Mismatches 62:	Indels 71:	Gaps 6
QY	5	AETTRDSFETTSGGSSVDLVGMALPGLAPNLSSASVSASASSEDASAKIRKPTTITSRGS	64	
DB	14	ATTTSDATATTT-----TATTGAGKEPEKKVRAATITTSRS	50	



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:03:27 ; Search time 109.58 seconds  
(without alignments)  
905.809 Million cell updates/sec

Title: US-09-640-211A-1076  
Perfect score: 1492  
Sequence: 1 MPMLAETRYRDSFETTSGCCS.....EWIQLIKLCTHPEPHNL 282

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues  
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/2/paa/PCTUS.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082.COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083.COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084.COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085.COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086.COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087.COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088.COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089.COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090.COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091.COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092.COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093.COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094.COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095.COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096.COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097.COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098.COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099.COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US100.COMB.pep:\*  
25: /cgn2\_6/ptodata/2/paa/US101.COMB.pep:\*  
26: /cgn2\_6/ptodata/2/paa/US60.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1492	100.0	282	1	PCT-US00-06112-1076
2	1492	100.0	282	20	US-09-640-211A-1076
3	724.5	48.6	290	18	US-09-452-997-12
4	701	47.0	332	26	US-60-324-109-23746
5	700	46.9	325	26	US-60-312-544-10669
6	695	46.6	327	1	PCT-US00-06112-2276
7	695	46.6	327	16	US-09-266-513-346

8	695	46.6	327	20	US-09-640-211A-2276	Sequence 2276, Ap
9	678	45.4	286	18	US-09-452-997-20	Sequence 20, Appl
10	664	44.5	317	26	US-60-312-544-10279	Sequence 10279, A
11	645	43.2	331	19	US-09-595-329A-2123	Sequence 2123, Ap
12	645	43.2	356	16	US-09-595-329A-2122	Sequence 2122, Ap
13	630	42.2	316	19	US-09-595-329A-2124	Sequence 2124, Ap
14	625	41.9	290	21	US-09-708-427-53221	Sequence 53221, A
15	625	41.9	295	21	US-09-708-427-53220	Sequence 53220, A
16	622.5	41.7	334	26	US-60-356-051-2044	Sequence 2044, Ap
17	608.5	40.8	302	20	US-09-620-394B-3147	Sequence 3147, Ap
18	606.5	40.7	284	20	US-09-620-394B-3148	Sequence 3148, Ap
19	598.5	40.1	302	1	PCT-US00-31414-12	Sequence 12, Appl
20	576	38.6	287	26	US-60-356-051-1779	Sequence 1779, Ap
21	574.5	38.5	293	17	PCT-US00-31414-44	Sequence 44, Appl
22	574.5	38.5	293	17	US-09-394-519-50	Sequence 50, Appl
23	574.5	38.5	293	26	US-60-312-544-7855	Sequence 7855, Ap
24	574.5	38.5	293	26	US-60-324-109-33179	Sequence 33179, A
25	567.5	38.0	307	26	US-60-356-051-2104	Sequence 2104, Ap
26	560.5	37.6	214	21	US-09-733-089-18848	Sequence 18848, A
27	560.5	37.6	214	22	US-09-816-660-18848	Sequence 18848, A
28	555	37.2	275	18	US-09-452-997-24	Sequence 24, Appl
29	551	36.9	422	21	US-09-733-089-761	Sequence 761, App
30	551	36.9	422	22	US-09-816-660-761	Sequence 761, App
31	521	34.9	264	21	US-09-733-089-14811	Sequence 14811, A
32	521	34.9	264	22	US-09-816-660-14811	Sequence 14811, A
33	520.5	34.9	278	26	US-60-312-544-8983	Sequence 8983, Ap
34	519.5	34.8	229	26	US-60-324-109-27074	Sequence 27074, A
35	519.5	34.8	229	26	US-60-324-109-33107	Sequence 33107, A
36	503.5	33.7	250	21	US-09-733-089-14750	Sequence 14750, A
37	503.5	33.7	250	22	US-09-816-660-14750	Sequence 14750, A
38	502.5	33.7	213	18	US-09-452-997-25	Sequence 25, Appl
39	502.5	33.7	213	21	US-09-708-427-8611	Sequence 8611, A
40	500.5	33.5	195	21	US-09-708-427-8612	Sequence 8612, Ap
41	497	33.3	129	21	US-09-733-089-22573	Sequence 22573, A
42	497	33.3	129	22	US-09-816-660-22573	Sequence 22573, A
43	492	33.0	130	21	US-09-708-427-51138	Sequence 51138, A
44	492	33.0	165	21	US-09-708-427-51136	Sequence 51136, A
45	490	32.8	128	26	US-60-312-544-6419	Sequence 6419, Ap

## ALIGNMENTS

RESULT 1  
PCT-US00-06112-1076  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Marion  
; APPLICANT: Shenk, Michael A.  
; APPLICANT: McGrath, Annette  
; TITLE OF INVENTION: Compositions and methods for the  
; FILE REFERENCE: 11000.1021C1PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/06112  
; NUMBER OF SEQ ID NOS: 2368  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 1076  
; LENGTH: 282  
; TYPE: PRT  
; ORGANISM: Pinus radiata  
PCT-US00-06112-1076

Query Match 100.0%; Score 1492; DB 1; Length 282;  
Best Local Similarity 100.0%; Pred. No. 6.3e-120;  
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 MPMLAETRYRDSFETTSGCCSVLDVGNALPGLAPNLSASVSASASDSAKKIRKPTTK 60  
|||||  
Db 1 MPMLAETRYRDSFETTSGCCSVLDVGNALPGLAPNLSASVSASASDSAKKIRKPTTK 60

```

QY 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVP 120
; SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVP 120
Db 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVP 120
QY 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAPPTATQLDSCYTPRAESSILITKSGSSCPT 100
; PPRKRRASHHPYPOKASKNVPVSOQVSTAPPTATQLDSCYTPRAESSILITKSGSSCPT 100
Db 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAPPTATQLDSCYTPRAESSILITKSGSSCPT 100
QY 181 VSSWVHHTTIFSIDASFVEKDDGPGPIETGNNCSGSGSTESSPTWPCSEIPEKVKPDFS 240
; VSSWVHHTTIFSIDASFVEKDDGPGPIETGNNCSGSGSTESSPTWPCSEIPEKVKPDFS 240
Db 181 VSSWVHHTTIFSIDASFVEKDDGPGPIETGNNCSGSGSTESSPTWPCSEIPEKVKPDFS 240
QY 241 QVYKFGSVDPSTTDHLKRLKEWIOILKLCCTHEEPFNL 282
; QVYKFGSVDPSTTDHLKRLKEWIOILKLCCTHEEPFNL 282
Db 241 QVYKFGSVDPSTTDHLKRLKEWIOILKLCCTHEEPFNL 282

```

```

RESULT 2
US-09-640-211a-1076
; Sequence 1076, Application US/09640211a
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; FILE REFERENCE: 11000.1021CIU
; CURRENT APPLICATION NUMBER: US/09/640,211a
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1076
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211a-1076

```

```

Query Match 100.0%; Score 1492; DB 20; Length 282;
Best Local Similarity 100.0%; Pred. No. 6.3e-120;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPMLAETTYRDSFETTSQSSVDYLGMLPGIAPRLSASVSASASEDSAKKIRPYITK 60
; MPMLAETTYRDSFETTSQSSVDYLGMLPGIAPRLSASVSASASEDSAKKIRPYITK 60
Db 1 MPMLAETTYRDSFETTSQSSVDYLGMLPGIAPRLSASVSASASEDSAKKIRPYITK 60
QY 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVP 120
; SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVP 120
Db 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVP 120
QY 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAPPTATQLDSCYTPRAESSILITKSGSSCPT 180
; PPRKRRASHHPYPOKASKNVPVSOQVSTAPPTATQLDSCYTPRAESSILITKSGSSCPT 180
Db 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAPPTATQLDSCYTPRAESSILITKSGSSCPT 180
QY 181 VSSWVHHTTIFSIDASFVEKDDGPGPIETGNNCSGSGSTESSPTWPCSEIPEKVKPDFS 240
; VSSWVHHTTIFSIDASFVEKDDGPGPIETGNNCSGSGSTESSPTWPCSEIPEKVKPDFS 240
Db 181 VSSWVHHTTIFSIDASFVEKDDGPGPIETGNNCSGSGSTESSPTWPCSEIPEKVKPDFS 240
QY 241 QVYKFGSVDPSTTDHLKRLKEWIOILKLCCTHEEPFNL 282
; QVYKFGSVDPSTTDHLKRLKEWIOILKLCCTHEEPFNL 282
Db 241 QVYKFGSVDPSTTDHLKRLKEWIOILKLCCTHEEPFNL 282

```

```

RESULT 3
US-09-452-997-12
; Sequence 12, Application US/09452997
; GENERAL INFORMATION:
; APPLICANT: Weng, Zude
; APPLICANT: Orozco, Emil M.
; TITLE OF INVENTION: Plant Circadian Rhythm-Myb Homologs
; FILE REFERENCE: B81296 US NA
; CURRENT APPLICATION NUMBER: US/09/452,997

```

```

; CURRENT FILING DATE: 1999-12-02
; EARLIER APPLICATION NUMBER: 60/110,780
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-452-997-12

```

```

Query Match 48.6%; Score 724.5; DB 18; Length 290;
Best Local Similarity 61.8%; Pred. No. 1.1e-53;
Matches 147; Conservative 27; Mismatches 37; Indels 27; Gaps 5;

```

```

QY 33 PNLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQLEFDRDKKIEAFVGS 92
; PNLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQLEFDRDKKIEAFVGS 92
Db 8 PPDAAAAGSAGEDASKVKVRKPYITTKRESMSQEHDKFLEALQLEFDRDKKIEAFVGS 67
QY 93 SKTVIQRSHAQKYFLKVKNGTREHVPVPPRKRASHHPYPOKASKNVPVSOQVSTAPPT 152
; SKTVIQRSHAQKYFLKVKNGTSEHVPVPPRKRASHHPYPOKASKNVPVSOQVSTAPPT 116
Db 68 SKTVIQRSHAQKYFLKVKNGTSEHVPVPPRKRASHHPYPOKASKNVPVSOQVSTAPPT 116
QY 153 AATQLDSCYTPRAESSILITKSGSSCPTVSSWVHHTTIFSIDASFVEKDDGPGPIETGNN 212
; AATQLDSCYTPRAESSILITKSGSSCPTVSSWVHHTTIFSIDASFVEKDDGPGPIETGNN 168
Db 117 -----GYTIKADSSMLRNSGMA-TVSSWTHNSIPPIVASSWYKEDLGAGANAPNF 168
QY 213 CSSGSTRESSPTWPCSEIPEKVKPDFS 263
; CSSGSTRESSPTWPCSEIPEKVKPDFS 263
Db 169 CSS-STEGPARANOPGETNDQINQVSLKMLMPDFAOVYFLSGLVDPSTISGLKQLE 225

```

```

RESULT 4
US-60-324-109-23746
; Sequence 23746, Application US/60324109
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)B
; CURRENT APPLICATION NUMBER: US/60/324,109
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 33196
; SEQ ID NO 23746
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
US-60-324-109-23746

```

```

Query Match 47.0%; Score 701; DB 26; Length 332;
Best Local Similarity 58.6%; Pred. No. 1.3e-51;
Matches 153; Conservative 24; Mismatches 58; Indels 26; Gaps 7;

```

```

QY 22 DLYGMLPG-----LAPLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQ 78
; DLYGMLPG-----LAPLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQ 75
Db 16 DPSNMTLPGVNNLPPPPRPAAPAAVAVEDPNKKIRPYITTKRESMSQEHDKFLEALQ 75
QY 79 LFDRODKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTREHVPVPPRKRASHHPYPOKASK 138
; LFDRODKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTSEHVPVPPRKRASHHPYPOKASK 135
Db 76 LFDRODKKIEAFVGSKTVIQIRSHAQKYFLKVKNGTSEHVPVPPRKRASHHPYPOKASK 135
QY 139 NNPVSO-----QVSTAFPPAATQLDSCYTPRAESSILITKSGSSCPTVSSWVHHTTIPST- 192
; NNPVSO-----QVSTAFPPAATQLDSCYTPRAESSILITKSGSSCPTVSSWVHHTTIPST- 188
Db 136 TPTVSQVMGRLSSSAF-----IEPAYIYSPDSSVLCTPTTNMP-LSSWYNTTPOG 188
; TPTVSQVMGRLSSSAF-----IEPAYIYSPDSSVLCTPTTNMP-LSSWYNTTPOG 242
QY 193 DASFVEKDDGPGPIETG---NNCSGSTRESSPTWPCSEI-----PEKVKPDFSQV 242

```



```
Db 189 NVPOVTRDDMDGLTGAGGAAPLNCYSSSSNESTPPTWPRSKRINOOGDKPIKVMPPFAQY 248
QY 243 YKFIGSVFDPSTDLKLKE 263
Db 249 YSFIGSVFDPSTNHLQLOO 269
```

```
RESULT 5
US-60-312-544-10669
; Sequence 10669, Application US/60312544
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Stein, Joshua
; APPLICANT: Liu, Jiongong
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)A
; CURRENT APPLICATION NUMBER: US/60/312, 544
; CURRENT FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 10730
; SEQ ID NO 10669
; LENGTH: 325
; TYPE: PRF
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: uC_gmromlsoy104a03_FLI
; US-60-312-544-10669
```

```
Query Match 46.9%; Score 700; DB 26; Length 325;
Best Local Similarity 57.9%; Pred. No. 1.7e-51;
Matches 143; Conservative 35; Mismatches 49; Indels 20; Gaps 5;
```

```
QY 25 GMAIPGLAPLNL-SASVSASASESDAKIRKPYTTKSRRESWEOEHDKFLEALQLPDRD 83
Db 43 GMPLEGPMPRASVSAATNAISPDEAAKIRKPYTTKSRRESWTEPHDKLEALQLPDRD 102
QY 84 WKTEAFVSGKTVIQRSHAKYFLKVKQNGREHVPPRRKASHPYPOKASKNVPS 143
Db 103 WKTEAFVSGKSVIQRSHAKYFLKVKQNGSTSEHLPPRRKRAHPYPOKASKNAPVL 162
QY 144 QQVSTAFPAATQDSCGYTTRAESSTLTKSGSCPTVSSVHHTIPSIDASFEYKDDG 203
Db 163 SQVSGRFOSSALIEPGYILKNAPPL-KTPIAMTVASSMSNLTQTANLS----- 213
QY 204 PGCIETGNCCSGSTESSPTWPCSE-----IPEKVPDFSOYKFIGSVFDPSTTD 256
Db 214 -PVTKVNNPCSSG--ESTPKVMPVSGESGOGNKTHPLRLVLPFTQYVGFISVFPDPNATE 270
QY 257 HLKRLKE 263
Db 271 HLKRLK 277
```

```
RESULT 6
PCT-US00-06112-2276
; Sequence 2276, Application PC/TUS0006112
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and methods for the
; FILE REFERENCE: 11000.1021C1PCT
; CURRENT APPLICATION NUMBER: PCT/US00/06112
; CURRENT FILING DATE: 2000-03-13
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2276
```

```
; LENGTH: 327
; TYPE: PRF
; ORGANISM: Eucalyptus grandis
; PCT-US00-06112-2276
```

```
Query Match 46.6%; Score 695; DB 1; Length 327;
Best Local Similarity 56.4%; Pred. No. 4.7e-51;
Matches 145; Conservative 33; Mismatches 55; Indels 24; Gaps 7;
```

```
QY 34 NLSASVSASA-----SEDSAKKIRKPYTTKSRRESWEOEHDKFLEALQLPDRDKKIEA 89
Db 26 NAGSAABEGGAAPVADSDSKVKRKYTTKSRRESWTEBDEHDKFLEALQLPDRDKKIEA 85
QY 90 FVSGKTVIQRSHAKYFLKVKQNGREHVPPRRKASHPYPOKASNNVPSQOVSTA 149
Db 86 FVSGKTVIQRSHAKYFLKVKQNGTSEHVPPRRKRAHPYPOKAPK-APVVSQVNP 144
QY 150 FPTAATQDSCGYTTRAESSTLTKSGSCPTVSSVHHTIPSIDASFEYKDDG--GPGI 207
Db 145 FOVSSAFLEPGHIVRPGSALGNSTV-AUSSMHSNVPMASASQGTCKDVGISGP-- 201
QY 208 ETGNCCSGSTESSPTWPCSEIPE-----KVPKDFSOYKFIGSVFDPSTDLK 260
Db 202 VPSNCCNCCSSNDSTPRSWNAQAIETPLDQKHLRVMPDFAQYVRFISGVFDPDAGHQLR 261
QY 261 LK-----EWIQLIK 270
Db 262 LKQMDPINLETVLAK 278
```

```
RESULT 7
US-09-266-513-346
; Sequence 346, Application US/09266513
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: McGrath, Annette
; APPLICANT: Shenk, M. Andrew
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions Isolated From Plant Cells
; TITLE OF INVENTION: and their use in the modification of Plant Cell
; FILE REFERENCE: 11000/1021
; CURRENT APPLICATION NUMBER: US/09/266, 513
; CURRENT FILING DATE: 1999-03-11
; NUMBER OF SEQ ID NOS: 438
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 346
; LENGTH: 327
; TYPE: PRF
; ORGANISM: Eucalyptus grandis
; US-09-266-513-346
```

```
Query Match 46.6%; Score 695; DB 16; Length 327;
Best Local Similarity 56.4%; Pred. No. 4.7e-51;
Matches 145; Conservative 33; Mismatches 55; Indels 24; Gaps 7;
```

```
QY 34 NLSASVSASA-----SEDSAKKIRKPYTTKSRRESWEOEHDKFLEALQLPDRDKKIEA 89
Db 26 NAGSAABEGGAAPVADSDSKVKRKYTTKSRRESWTEBDEHDKFLEALQLPDRDKKIEA 85
QY 90 FVSGKTVIQRSHAKYFLKVKQNGREHVPPRRKASHPYPOKASNNVPSQOVSTA 149
Db 86 FVSGKTVIQRSHAKYFLKVKQNGTSEHVPPRRKRAHPYPOKAPK-APVVSQVNP 144
QY 150 FPTAATQDSCGYTTRAESSTLTKSGSCPTVSSVHHTIPSIDASFEYKDDG--GPGI 207
Db 145 FOVSSAFLEPGHIVRPGSALGNSTV-AUSSMHSNVPMASASQGTCKDVGISGP-- 201
QY 208 ETGNCCSGSTESSPTWPCSEIPE-----KVPKDFSOYKFIGSVFDPSTDLK 260
Db 202 VPSNCCNCCSSNDSTPRSWNAQAIETPLDQKHLRVMPDFAQYVRFISGVFDPDAGHQLR 261
```

```
QY      261 LK-----EWIQLLK 270
          ||      | : | : |
Db      262 LKQMPINLETVLLMK 278
```

RESULT 8  
US-09-640-211A-2276

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match:	46.68:	Score 695:	DB 20:	Length 327:
Best Local Similarity:	56.48:	Pred. No. 4.7e-51:		
Matches 145:	Conservative 33:	Mismatches 55:	Indels 24:	Gaps 7:

[illegible]

```

RESULT 9
US-09-452-997-20
: Sequence 20 Application US/09452997
: GENERAL INFORMATION:
: APPLICANT: Weng, Zude
: APPLICANT: Orozco, Emil M.
: TITLE OF INVENTION: Plant Circadian Rhythm-Myb Homologs
: FILE REFERENCE: B81996 US NA
: CURRENT APPLICATION NUMBER: US/09/452,997
: CURRENT FILING DATE: 1999-12-02
: EARLIER APPLICATION NUMBER: 60/110,780
: EARLIER FILING DATE: 1998-12-03
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 20
: LENGTH: 286
: TYPE: PRF
: ORGANISM: Triticum aestivum
: US-09-452-997-20

```

Query Match	45.48;	Score 678;	DB 18;	Length 286;
Best Local Similarity	57.98;	Pred. No. 1.2e-49;		
Matches 140;	Conservative 27;	Mismatches 47;	Indels 28;	Gaps 6

[illegible]

```

RESULT 10
US-60-312-544-10279
: Sequence 10279, Application US/60312544
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Edgerton, Michael D
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Stein, Joshua
: TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
: FILE REFERENCE: 38-10(52726)A
: CURRENT APPLICATION NUMBER: US/60/312,544
: CURRENT FILING DATE: 2001-08-15
: NUMBER OF SEQ ID NOS: 10730
: SEQ ID NO 10279
: LENGTH: 317
: TYPE: PRF
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB3242-470-B8_FLI
US-60-312-544-10279

```

[illegible]

```

; LOCATION (1)..(356)
; OTHER INFORMATION: Ceres Seq. ID no. 1015308
; NAME/KEY: misc_feature
; LOCATION: ()..(1)
; OTHER INFORMATION: Xaa is any aa, unknown or other
; US-09-595-329A-2122

Query Match          43.28; Score 645; DB 19; Length 356;
Best Local Similarity 52.7%; Pred. No. 1,1e-46;
Matches 147; Conservative 31; Mismatches 55; Indels 46; Gaps

QY 16 SCGSSVDVLVGMALBGLAPMLNSASVSASAS-----EDSAKTRK 54
      | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
Db 31 SDGFLDPPGMYPLVGLGPFsf-TAAVSSSSSPPTSTTAAVADVATAMVSSSEDSLKKTR 89
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 55 PYRTTKRESMSQEODHKFLKALQLFDRDMKKIEAFVGSKTVIOIRSHAQKFLPKVONG 114
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 PYRTTKRESMTPEPHDKFLKALQLFDRDMKKIEAFISKTVIOIRSHAQKFLPKVONGK 149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 115 TREHVPPEPRKKRASHPRPOKASKNNVPSQVSTAFPTAATQLDSCGYVPARESSSILTKS 174
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 TGEHLPEPRKKRAAHPYPOKAHKNNQVL-QVPGSFKSTSEPNDFPFMFRESSMLMTS 207
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 175 GSSCPYVSSMWVHTFISIDASFYEKDDGCPPEIETGNCCSGSTIESSPPTWPPCSELPF- 233
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 208 ----PTTAAALPPTNNAOTISFTPL-----PKGAGANNCCSS-SSSNP-----RPSNRDAR 255
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

```

RESULT 13
US-09-595-329A-2124
: Sequence 2124, Application US/09595329A
: GENERAL INFORMATION:
: APPLICANT: ALEXANDROV, Nickolai
: APPLICANT: VYACHESLAV, Brover
: TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypep
: TITLE OF INVENTION: Thereby
: PTE REFERENCE: 2750-0948P
: CURRENT APPLICATION NUMBER: US/09/595.329A
: CURRENT FILING DATE: 2000-06-16
: NUMBER OF SEQ ID NOS: 3309
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 2124
: LENGTH: 316
: TYPE: PRT
: ORGANISM: Arabidopsis Thaliana
: FEATURE:
: NAME/KEY: peptide
:

```

```

: NAME/KEY misc:feature
: LOCATION: {}-{}
: OTHER INFORMATION: Xaa is any aa, unknown or other
US-09-595-329A-2124

Query Match          42.2%; Score 630; DB 19; Length 316;
Best Local Similarity 53.2%; Pred. No. 1,8e+45;
Matches 14; Conservative 30; Mismatches 50; Indels 46; Gaps

QY 26 MALPGLAPNLSASVSASAS-----EDSAKKIRKRYTTTKRES 64
      |::||:|:|::||:|:|      ||:|||||
Db 1  MTVGGLGPSF-TAAVSSSSPTTSTVAVAADVATAMVSSSEEDLSKTKRKRYTTTKRES 59
      |::||:|:|::||:|:|      ||:|||||

QY 65 MSDEHHRKFLLEALQLEPRDMKITEAVGVSKTYIQRSHAKQFLKLVQNGRREHPVPPR 124
      |::||:|:|::||:|:|      ||:|||||
Db 60 WTEDEHHRKFLLEALQLEPRDMKITEAVIGSKTYIQRSHAKQFLKLVQNGSGEHLPPPPR 119
      |::||:|:|::||:|:|      ||:|||||

QY 125 KRKASHPEPKASKNNVSVQAVSTAEPATATOLDGCVYPRAESSILTKSGSSCPTVSW 184

```

```

Db 120 KRKAHPYPOAHKNVOL--QVPSFSTSEPNPSPMPRESSMLMTS---PTTAA 173
QY 185 VHTIPSIDSFVEKDDGPGIETGNSSGSGTSSPTWPCSEIPE-----K 234
Db 174 APMTNNMOTISFTL-----PKGAGANNCS--SENTP---RRSNDARDHGVGSLR 225
QY 235 VKPDFSOVKFISGVDPSTTDHLKLE 263
Db 226 VLPDFAQVYGFISGVDFPYASNLQKLK 254

```

RESULT 14  
US-09-708-427-53221

; Sequence 53221, Application US/09708427

; GENERAL INFORMATION:

; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE OF INVENTION: THEREBY

; FILE REFERENCE: 2750-1243P

; CURRENT APPLICATION NUMBER: US/09/708,427

; NUMBER OF SEQ. ID NOS: 85364

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 53221

; LENGTH: 290

; TYPE: PRT

; ORGANISM: zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1..290

; OTHER INFORMATION: Xaa is any amino acid

; NAME/KEY: misc\_feature

; LOCATION: 1..290

; OTHER INFORMATION: Ceres Seq. ID 1930758

; US-09-708-427-53221

Query Match 41.9%; Score 625; DB 21; Length 290;  
Best Local Similarity 53.6%; Pred. No. 4.4e-45;  
Matches 135; Conservative 23; Mismatches 50; Indels 44; Gaps 8;

```

QY 26 MALGLAPNLSASVSASASDSAKKIRKPYTIKRSRESSEQHDKFLALQLFDRDK 85
Db 1 MVAAPAA-----AAEEYRKVKRPYTIKRSRESWTEPHDKFLALQLFDRDK 50
QY 86 KIEFAVSKTYIQRSHAQKFLVKQNGTREHVPKPKKASHPPYQKSKNVP--VSO 144
Db 51 KIEFAVSKTYIQRSHAQKFLVKQNGTGEHLPPKPKKAAHPYQKSKNAPAVSQ 110
QY 145 QV-STAPPTAATQDLSGYYPRAESSILTKSGSSCPYVSSVWHT-----IPSIDASFE 198
Db 111 AILSEQPTQREQ-----GSVPMDDTATVRNTNANVAVPSMDNTLAQ 152
QY 199 KDDCG-PPGIETGNSSGSGTSSPTWPCSEI-----PEVKPDFSOVKFISGVF 250
Db 153 PFSAGHVOGAATNNCS--SMESPSGTFPTSEAVEQENMVPPLRAMPDFAQVYVNFISIF 211
QY 251 DPSTTDHLKLE 262
Db 212 DPSTTDHLKLE 223

```

RESULT 15

US-09-708-427-53220

; Sequence 53220, Application US/09708427

; GENERAL INFORMATION:

; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE OF INVENTION: THEREBY

; FILE REFERENCE: 2750-1243P

; CURRENT APPLICATION NUMBER: US/09/708,427

; CURRENT FILING DATE: 2000-11-09

```

; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53220
; LENGTH: 295
; TYPE: PRT
; ORGANISM: zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..295
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..295
; OTHER INFORMATION: Ceres Seq. ID 1930757
; US-09-708-427-53220

```

Query Match 41.9%; Score 625; DB 21; Length 295;  
Best Local Similarity 54.3%; Pred. No. 4.5e-45;  
Matches 134; Conservative 24; Mismatches 55; Indels 34; Gaps 7;

```

QY 31 LAFNLSASVSASASDSAKKIRKPYTIKRSRESSEQHDKFLALQLFDRDKKIEAF 90
Db 1 MAPPPVVPAAAPAAEEYRKVKRPYTIKRSRESWTEPHDKFLALQLFDRDKKIEAF 60
QY 91 VGSKTIVYIQRSHAQKFLVKQNGTREHVPKPKKASHPPYQKSKNVP--VSOQV-ST 148
Db 61 VGSKTIVYIQRSHAQKFLVKQNGTGEHLPPKPKKAAHPYQKSKNAPAVSQALISQ 120
QY 149 APTTATQDLSGYYPRAESSILTKSGSSCPYVSSVWHT-----IPSIDASFEKDDCG 203
Db 121 EOPTQREQ-----GSVPMDDTATVRNTNANVAVPSMDNTLAQPFSAQ 162
QY 204 -PPGIETGNSSGSGTSSPTWPCSEI-----PEVKPDFSOVKFISGVDPSTT 255
Db 163 HVGGAATNNCS--SMESPSGTFPTSEAVEQENMVPPLRAMPDFAQVYVNFISIFDPOTS 221
QY 256 DHTLKLK 262
Db 222 GHTLQMLK 228

```

Search completed: August 12, 2002, 10:10:41  
Job time: 434 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:03:02 ; Search time 13.08 Seconds

(Without alignments)  
526.607 Million cell updates/sec

Title: US-09-640-211a-1076

Perfect score: 1492  
Sequence: 1 MPMLAEFRDSEFTSGSS.....EWIQLILKCTHEEPFHL 282

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCFUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	335	22.5	83	US-09-202-712-16	Sequence 16, Appl
2	322	21.6	645	US-09-202-712-2	Sequence 2, Appl
3	238	16.0	83	US-09-202-712-14	Sequence 14, Appl
4	200	13.4	57	US-09-202-712-13	Sequence 13, Appl
5	195	13.1	55	US-09-202-712-7	Sequence 7, Appl
6	107	7.2	1123	US-08-408-865-1	Sequence 1, Appl
7	100.5	6.7	1105	US-08-999-774A-2	Sequence 2, Appl
8	96	6.4	1958	US-07-945-283-2	Sequence 2, Appl
9	92.5	6.2	568	US-08-320-559-30	Sequence 30, Appl
10	92.5	6.2	568	US-08-545-860D-30	Sequence 30, Appl
11	92.5	6.2	568	PCT-US94-004496-30	Sequence 30, Appl
12	91.5	6.1	410	US-09-431-573-4	Sequence 4, Appl
13	91.5	6.1	456	US-08-819-013-1	Sequence 1, Appl
14	91.5	6.1	1005	US-08-935-450-2	Sequence 2, Appl
15	90.5	6.1	410	US-08-123-343A-5	Sequence 5, Appl
16	90.5	6.1	410	US-08-123-343A-7	Sequence 7, Appl
17	90	6.0	430	US-08-427-993B-7	Sequence 7, Appl
18	90	6.0	430	US-08-478-609A-7	Sequence 7, Appl
19	90	6.0	509	US-08-427-993B-1	Sequence 1, Appl
20	90	6.0	509	US-08-478-609A-1	Sequence 1, Appl
21	90	6.0	1185	US-09-041-886-23	Sequence 23, Appl
22	87.5	5.9	268	US-09-187-049-9	Sequence 9, Appl
23	87.5	5.9	1259	US-09-187-049-13	Sequence 13, Appl
24	87	5.8	369	US-08-773-870-4	Sequence 4, Appl
25	87	5.8	369	US-09-393-569-2	Sequence 2, Appl
26	86.5	5.8	410	US-09-431-573-5	Sequence 5, Appl
27	86.5	5.8	672	US-08-049-254-2	Sequence 2, Appl

28	86.5	5.8	672	1	US-08-472-934-2	Sequence 2, Appl
29	86.5	5.8	672	2	US-08-323-460A-2	Sequence 2, Appl
30	86.5	5.8	672	2	US-08-461-146C-2	Sequence 2, Appl
31	86.5	5.8	672	3	US-08-461-145C-2	Sequence 2, Appl
32	86.5	5.8	672	4	US-08-628-829-2	Sequence 2, Appl
33	86.5	5.8	793	4	US-09-588-256-10	Sequence 10, Appl
34	86.5	5.8	1302	4	US-09-423-890-2	Sequence 2, Appl
35	86.5	5.8	1493	4	US-08-423-890-8	Sequence 8, Appl
36	86.5	5.8	1593	4	US-08-628-829-4	Sequence 4, Appl
37	86	5.8	553	3	US-09-083-351-2	Sequence 2, Appl
38	86	5.8	553	4	US-09-083-352-2	Sequence 2, Appl
39	86	5.8	1627	1	US-07-665-792E-9	Sequence 9, Appl
40	86	5.8	2842	1	US-07-741-940-7	Sequence 7, Appl
41	86	5.8	2842	1	US-08-289-548A-7	Sequence 7, Appl
42	86	5.8	2842	1	US-08-452-654-7	Sequence 7, Appl
43	86	5.8	2843	1	US-07-741-940-2	Sequence 2, Appl
44	86	5.8	2843	1	US-08-289-548A-2	Sequence 2, Appl
45	86	5.8	2843	1	US-08-452-654-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-202-712-16
; Sequence 16, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffer, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 16
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-202-712-16

Query Match 22.5%; Score 335; DB 4; Length 83;
Best Local Similarity 80.0%; Pred. No. 2,9e+24;
Matches 64; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 41 SASASEDSAKIRKPYTTTSRESWSQEHDKFLBALQFLDRDWKKIEAFVSGKTYIQIR 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 TTGEAPEKKVRRAYITTSRESWTEGHEHDKFLBALQFLDRDWKKIEAFVSGKTYIQIR 62

QY 101 SHAOXFLYKOKNGTREHVP 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 SHAOXFLYKOKNGTIAHP 82

RESULT 2
US-09-202-712-2
; Sequence 2, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffer, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23

```



SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1123  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-408-865-1

Query Match 7.2%; Score 107; DB 4; Length 1123;  
Best Local Similarity 27.3%; Pred. No. 0.22; Indels 56; Gaps 12;  
Matches 65; Conservative 25; Mismatches 92; Indels 56; Gaps 12;

QY 36 SSASVASASASDSAKIRKRYT-----ITKSRE-SMSEQEHDKFLALQLEFRDMKRIEAF 90  
DB 611 SSSSPHSASDSSTKAPQTPRSAAHLCDSEINCSLAGSKRPISGA--DSKTVLAKSP 668  
QY 91 VSKTVIQRIS-----HAQYFLKVKOK-----NCTREHVPPRRKRASHPYPOKASK 138  
DB 669 VLSNTTTPASTMSPPKALALSAKASTYLMRATGNDLRRPPSPSSDLTH--PKTSH 726  
QY 139 NVPVSOQVSTAPPTATQDLSGYPRAESLITKSGSSCPYSSMVHTIPSIDASFYB 198  
DB 727 PV-----VASTWVHRRAVVS---PAPQSSRLQPPSPSPHTLLS---STPK----- 767  
QY 199 KDDGCPGIEGTGNCSGSTE-----SSPTWPCSEIPEKYPFDSQYKPF 247  
DB 768 -----PPGTSEPRSSCSISITALPQVNEDLVSLPHQLEPASERPQ--SPSEKRRKKTIVG 818

RESULT 7  
US-08-999-774A-2  
Sequence 2, Application US/08999774A  
Patent No. 6274312

GENERAL INFORMATION:  
APPLICANT: Gish, Kurt C.  
APPLICANT: Seghezzi, Wolfgang  
APPLICANT: Shanahan, Frances  
APPLICANT: Lees, Emma M.  
APPLICANT: McLanahan, Terrell K.  
TITLE OF INVENTION: Intracellular Regulatory Molecules;  
TITLE OF INVENTION: Related Reagents  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/999,774A  
FILING DATE: 10-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/032,818  
FILING DATE: 11-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0646  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650)852-9196  
TELEFAX: (650)496-1200  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1105 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-999-774A-2

Query Match 6.7%; Score 100.5; DB 4; Length 1105;  
Best Local Similarity 22.3%; Pred. No. 0.87;  
Matches 59; Conservative 31; Mismatches 88; Indels 87; Gaps 12;

QY 30 GLAP-NLSASVASASAS-----EDSAK-----IKKPTITKSR-----ESMS 66  
DB 565 GIVPLHLRSPPQVPAQOMLNPEKKKREKRVLDQNGRLTDIYSKRTLAKSKASAGRWMT 624  
QY 67 EDEHOKFLALQLEFRDMKRIEAFGSKTVIQRSHAKYFLKVOKNGTREHVPPRRKR 126  
DB 625 EDETLLELLEALNKYDDMKVSEHVSGSRQDCILH-----FLRL----- 664  
QY 127 KASHPYPOK-----ASKNVPVSOQVSTAPPTATQDLSGYPRAESLITKSGSSC 178  
DB 665 PLEDPLYENSDASLGPAYQVPFSGSNPVMSTVAF--LASVDDRVASA-----AKAA 718  
QY 179 PTVSSVYVHTIP-----SIDASFYEKDDGGPPGIEGTGNCSGSTE 219  
DB 719 LEEFSRVREEVPLDELVAAHVKKVQEAARASGRVDEPTY-----GLE--SSCIATGCP 767  
QY 220 SSPTWPCSEIPEKYPFDSQYK 244  
DB 768 DEPEKLEGAEEKMEADPPDGOPEK 792

RESULT 8  
US-07-945-283-2  
Sequence 2, Application US/07945283  
Patent No. 5352596

GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
APPLICANT: Mesley, Ronald D.  
TITLE OF INVENTION: Pseudotables Virus Deletion Mutants  
TITLE OF INVENTION: Involving The EP0 and LIT Genes  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P.  
REGISTRATION NUMBER: 27976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 309-685-4011 ext.513  
TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1958 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-945-283-2

Query Match 6.4%; Score 96; DB 1; Length 1958;  
Best Local Similarity 26.8%; Pred. No. 5.1;  
Matches 41; Conservative 14; Mismatches 64; Indels 34; Gaps 7;

```

OY 110 VOKNTEHVNPPRRKRAKSAHRYOKAKSNRYVQOXYSTAFPRATQLODQY-----P 163
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 288 VQADQAVNDGR-----SSNSHGRGRGR-GRSTHQRPRGRPTSTNNNNNNHQRPTSP 342
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 164 RAESSILTKRGSSQPTVYSSWVNH-----TPISDAFVEKQ--DQGRGLETGNCS 214
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 343 RPTSSHODPRQGRPRSAETNNNNHODPRQSGGRPTSTNNNNHODPRQSGGRSPRRPSTS 402
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 215 SGSTESSPT-----WPPSETEPKV 235
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 403 SSSSHQGRPTSPRRQRPRRRPPPS--PQKI 433
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT      9
US-08-320-559-30
; Sequence 30, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
;

```

```

1  TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods for
2  TITLE OF INVENTION:  Detection and Treatment of Acute Leukemias
3  TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the
4  NUMBER OF INVENTIONS:  All-1 Region
5  NUMBER OF SEQUENCES:  44
6  CORRESPONDENCE ADDRESS:
7  ADDRESSSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 563335r1s
8  STREET:  One Liberty Place - 46th Floor
9  CITY:  Philadelphia
10 STATE:  PA
11 COUNTRY:  USA
12 ZIP:  19103
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 MB STORAGE
15 COMPUTER:  IBM PS/2
16 OPERATING SYSTEM:  PC-DOS
17 SOFTWARE:  WORDPERFECT 5.1
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/320,559
20 FILING DATE:
21 CLASSIFICATION:  435
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER:  US/08/062,443
24 FILING DATE:  14 MAY 1993
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:  US/07/971,094
27 FILING DATE:  30-OCT-92
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER:  US/07/888,830
30 FILING DATE:  27-MAY-92
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER:  US/07/805,093
33 FILING DATE:  11-DEC-91
34 ATTORNEY/AGENT INFORMATION:
35 NAME:  Deluca, Mark
36 REGISTRATION NUMBER:  33,229
37 REFERENCE/DOCKET NUMBER:  TJU-0855
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE:  (215) 568-3100
40 TELEFAX:  (215) 568-3439
41 INFORMATION FOR SEQ ID NO:  30:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH:  568 amino acids
44 TYPE:  amino acid
45 TOPOLOGY:  linear
46 MOLECULE TYPE:  protein
47 US-08-320-559-30

```

Query Match	6.2%;	Score 92.5;	DB 1;	Length 568;
Best Local Similarity	23.0%;	Pred. No. 1.9;		
Matches	59;	Conservative	34;	Mismatches 123;
			Indels	41;
			Gaps	8

[illegible]

```

10
US-08-545-860D-30
Sequence 30, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESSEE: No. 6040140r1s
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093

```







STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,343A  
FILING DATE: 17-SEP-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,957  
FILING DATE: 15-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5907A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 61861-9540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-123-343A-5

Query Match 6 1% Score 90.5; DB 1; Length 410;  
Best Local Similarity 24.0%; Pred. No. 1.9;  
Matches 59; Conservative 26; Mismatches 82; Indels 79; Gaps 13;  
QY 15 TSGGSSVDLVGMALPGLAPNLSASVSASASEDSARKI--RKPYITLTKSRESWSEOE--- 69  
DB 35 SSASSSVSSGVS-----SASASSASSSSSSSDGASSASQSPNTTSSATQTPMQSLP 90  
QY 70 HDKFLDALDLFDMDKKIEAFVGSKTIVQIRSHAOKYFLKVKNGTREHVPPPRK---- 125  
DB 91 TDQVLVLYALY---EMVR-----MYOSQGSAPQIF-----QYPPSPSCNFT 127  
QY 126 -----RKASHPYQKASKNYPVSG-----OVSTAPTAATQLDGYY 162  
DB 128 GGDFVFPFGHPNPNNSNHPRTPTSTVSFSSGEEYNFRRQOQPQHPSTYPAVSTP-----Q 182  
QY 163 PRAESSSILTKSGSGCPTVSSWVHTIPSIDASFVEKDDG-----PPGIEFGNNCS 215  
DB 183 PMRPOGAPRMHCHSHYPOGSA---HMMPHHSAPF---GMGGTYAGYTPP--PTPNTASA 234  
QY 216 GSTESS 221  
DB 235 GTSSSS 240

Search completed: August 12, 2002, 10:08:45  
Job time: 343 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 12, 2002, 10:02:37 : Search time 29.84 Seconds  
(without alignments)  
1049.692 Million cell updates/sec

Title: US-09-640-211a-1076  
Perfect score: 1492  
Sequence: 1 MPMLAEYRDSFETTSGGSS.....EMILILKLCCTHEEPFNL 282

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

1: A.Geneseq\_032802.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1492	100.0	282	21	AAAB33027
2	695	46.6	327	21	AAAB33321
3	631.5	42.3	333	21	AAAG36897
4	616.5	41.3	318	21	AAAG36898
5	608.5	40.8	302	21	AAAG45426
6	608	40.8	303	21	AAAG45426
7	606.5	40.7	284	21	AAAG45426
8	606	40.6	285	21	AAAG45427
9	598.5	40.1	284	21	AAAG36899
10	598.5	40.1	302	22	AAE02543
11	574.5	38.5	293	22	AAE02559

12	568.5	38.1	293	21	AAAG23940	Arabidopsis thalia
13	480.5	32.2	142	21	AAAB32797	Eucalyptus grandis
14	480.5	32.2	150	21	AAAB33245	Eucalyptus grandis
15	322	21.6	178	21	AAAG20007	Arabidopsis thalia
16	322	21.6	203	21	AAAG20006	Arabidopsis thalia
17	322	21.6	645	19	AAAG42084	Amino acid sequenc
18	322	21.6	645	22	AAE02547	A. thaliana transc
19	322	21.6	645	22	AAE01903	Arabidopsis thalia
20	322	21.6	645	22	AAE01926	Arabidopsis thalia
21	312	20.9	287	22	AAE01952	Arabidopsis thalia
22	310.5	20.8	836	21	AAAG42123	Arabidopsis thalia
23	310.5	20.8	856	21	AAAG42122	Arabidopsis thalia
24	310.5	20.8	909	21	AAAG42121	Arabidopsis thalia
25	307	20.6	163	21	AAAG20008	Arabidopsis thalia
26	305	20.4	608	19	AAAG79280	Arabidopsis thalia
27	305	20.4	608	22	AAE02563	A. thaliana transc
28	305	20.4	608	22	AAE01889	Arabidopsis thalia
29	305	20.4	608	22	AAE01911	Arabidopsis thalia
30	302	20.2	59	21	AAAB33396	Eucalyptus grandis
31	248.5	16.7	102	21	AAAB33278	Pinus radiata tran
32	239	16.0	47	21	AAAB33395	Eucalyptus grandis
33	207	13.9	145	21	AAAB33174	Eucalyptus grandis
34	168	11.3	48	21	AAAB33374	Pinus radiata tran
35	145	9.7	42	21	AAAB33375	Pinus radiata tran
36	144.5	9.7	309	21	AAAB33232	Eucalyptus grandis
37	144.5	9.7	370	21	AAAB32774	Arabidopsis thalia
38	137	9.2	387	21	AAAG49848	Arabidopsis thalia
39	136.5	9.1	203	21	AAAG46375	Arabidopsis thalia
40	136.5	9.1	215	21	AAAG46374	Arabidopsis thalia
41	133.5	8.9	275	21	AAAG08359	Arabidopsis thalia
42	133.5	8.9	278	21	AAAG08358	Arabidopsis thalia
43	133.5	8.9	314	21	AAAG08357	Arabidopsis thalia
44	133.5	8.9	322	21	AAAG49849	Arabidopsis thalia
45	133.5	8.9	384	21	AAAG19453	Arabidopsis thalia

## ALIGNMENTS

RESULT 1	
AAAB33027	standard; Protein; 282 AA.
XX	
AC	AAAB33027;
XX	
DT	25-JAN-2001 (first entry)
XX	
DE	Pinus radiata transcription factor protein sequence #154.
XX	
KW	Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW	poplar; sweetgum; teak; mahogany; bzip; G-box binding factor;
KW	basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW	homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;
XX	type 2 Cys2His2; CCAAT box element; MYB.
OS	Pinus radiata.
XX	
PN	WO200053724-A2.
XX	
PD	14-SEP-2000.
XX	
PF	09-MAR-2000; 2000MO-US06112.
XX	
PR	11-MAR-1999; 99US-0266513.
XX	
PA	18-AUG-1999; 99US-0149485.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	(FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX	
PI	Wood M, McGrath A, Shenk MA, Glenn M;
XX	
DR	WPI; 2000-579369/54.
XX	

PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT  
 XX  
 XX  
 PS Claim 8; Page 398; 747pp; English.  
 XX  
 CC The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
 CC transcription factor. The transcription factor may be used to produce a  
 CC plant having modified gene expression such as a woody plant e.g. a  
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
 CC to modify the activity of a polypeptide in a plant. The transcription  
 CC factors of the present invention are members from the following families  
 CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
 CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
 CC Cys2His2, CCAAT box elements and MYB.  
 CC  
 XX  
 SQ Sequence 282 AA;  
 Query Match 100.0%; Score 1492; DB 21; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 4,2e-129;  
 Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MPMLAETRYROSFTTSGSSVDLVGMALPGIAPMLSSASVSASASEDSAKKIRKPYITRK 60  
 Db 1 mpmiaelyrdrdfeltsgssvdilvgmalpgiapmlssasvsasasedsakkirkpytlk 60  
 OY 61 SRESMSDEHDKFLDALQLEDRDMKITEAFVGSKTVIQIRSHAKYFLKVKNGKTRHVP 120  
 Db 61 sresmsdehdkflleaqlldfdrdmkkieafvgsktvliqrshaqkyflkvngktrhvp 120  
 OY 121 PPRKAKRASHYPPKASKNNVPSQOVSTAFPTATOIDSQVTPRAESSITITKSGSCPT 180  
 Db 121 pprkkrashpyppkasknnvpsqvsatafptaaltqidsygytpraessitltksqsscpt 180  
 OY 181 VSSVNHHTIPSIDASFVEKDDGPGGIEGTGNCSSGSTESSPPTWPCSEIPEKVKRDFDS 240  
 Db 181 vssvnhhtipdsidfekddgpggietgncssgstesspwtwpcseipekvkrpfdfs 240  
 OY 241 QYVFRTGVSFDPSTTDLKRLKEWIOILIKICTHPEPFHNL 282  
 Db 241 qyvkftgsvfdpstdtlhklkewiqlilklctcheepfhnl 282  
 RESULT 2  
 AAB33321  
 ID AAB33321 standard; Protein: 327 AA.  
 XX  
 AC AAB33321;  
 XX  
 DT 25-JAN-2001 (first entry)  
 XX  
 DE Eucalyptus grandis transcription factor protein sequence #419.  
 XX  
 KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 PN WO200053724-A2.  
 PD 14-SEP-2000.  
 XX  
 PF 09-MAR-2000; 2000MO-US06112.  
 XX  
 PR 11-MAR-1999; 99US-0266513.  
 PR 18-AUG-1999; 99US-0149485.

XX  
 PA (GENE-) GENESIS RES 6 DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 PI Wood M, McGrath A, Shenk MA, Glenn M;  
 XX  
 DR WPI; 2000-579369/54.  
 XX  
 PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT  
 XX  
 XX  
 PS Claim 8; Page 724; 747pp; English.  
 XX  
 CC The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
 CC transcription factor. The transcription factor may be used to produce a  
 CC plant having modified gene expression such as a woody plant e.g. a  
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
 CC to modify the activity of a polypeptide in a plant. The transcription  
 CC factors of the present invention are members from the following families  
 CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
 CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
 CC Cys2His2, CCAAT box elements and MYB.  
 CC  
 XX  
 SQ Sequence 327 AA;  
 Query Match 46.6%; Score 695; DB 21; Length 327;  
 Best Local Similarity 56.4%; Pred. No. 1.4e-55;  
 Matches 145; Conservative 33; Mismatches 55; Indels 24; Gaps 7;  
 OY 34 NLSASVSASA-----SEDSAKKIRKPYITRKRESMSDEHDKFLDALQLEDRDMKITA 89  
 Db 26 nagsaaeggaapayaedskvkrkytlktsreswteghdflleaqlldfdwmkkia 85  
 OY 90 FVCSKTVIQIRSHAKYFLKVKNGKTRHVPPRPRKASHYPPKASKNNVPSQOVSTRA 149  
 Db 86 fvscktviqirshaqkyflkvngktrsehvpprprkrashpyppkasknnvpsqovst 144  
 OY 150 FPPAATOIDSQVTPRAESSITITKSGSCPTVSSVNHHTIPSIDASFVEKDDG--GPGCI 207  
 Db 145 fpyssaflpghlvprpdsallgnrstv-nlssvshnsvpamsaagtkdyisgpr-- 201  
 OY 208 ETGNMNCSSGSTESSPPTWPCSEIPE-----KVKPDSQVYKFTGVSFDPSTTDLK 260  
 Db 202 vpsncsssndstprswpnagaiepldqgkhlrvmpdtaqvryflgsvfdpdaaghlqr 261  
 OY 261 LK-----EWIOILIK 270  
 Db 262 lkqmdplnietvllmk 278  
 RESULT 3  
 AAG36897  
 ID AAG36897 standard; Protein: 333 AA.  
 XX  
 AC AAG36897;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 45282.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.

PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126764.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151330.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.

```
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161999.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161983.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 42.3%; Score 631.5; DB 21; Length 333;  
Best Local Similarity 51.0%; Pred. No. 1e-49;  
Matches 146; Conservative 31; Mismatches 60; Indels 49; Gaps 8;

```
QY 16 SCGSSVDLVGMALPGLAPNLSSASVSASAS-----EDSAKKIRK 54
D 6 sdgyfIdpTgtmTpyglpsf-taavsssspsltslavavadvamvsssedlsktkrk 64
QY 55 PYTTKRSRESSEQEHDKFLFALQFDRDMKKEAFVGSKTVIOIRSHAOKYFLKQKNG 114
D 65 pyltkstreswlepehdkflcaiqldfdwmkkaeflgsklcvldirshagkyflkvqksq 124
QY 115 TREHVPFRPRKASHPYOKASKNVPSQOVSTAFPTATQLDGYYPRAESSSILTKS 174
D 125 tgenlpprprrkkaahpyqkahnvq--qvpstkslaeapndpslmfpressmlmta 182
QY 175 GSSCPYVSWVNHHTIPSDASFVEKDDC-----GPPGIEGTGNCSSGSGSPPTWPP 227
D 183 ---pltaaaarwlnaqclstfclpkyccstlmetagagannccs-sseentp---rp 234
QY 228 CSEIPE-----KVKPDFSQVYKPIGSPDSTTDHLKKIKE 263
D 235 rsndardhngvshlrvlpdfaqvgflgsvldpyasnlnqlklk 280
```

```
RESULT 4
ID AAC36898 standard; Protein; 318 AA.
AC AAC36898;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45283.
```

```
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
```





```
RESULT      5
ID      AGI4908
XX      ACI4908 standard; Protein: 302 AA.
AC      ACI4908;
XX      17-OCT-2000 (first entry)
XX      Arabidopsis thaliana protein fragment SEQ ID NO: 14947.
XX      Protein identification: signal transduction pathway; metabolic pathway;
XX      KW      hybridisation assay; genetic mapping; gene expression control; promoter;
XX      OS      Arabidopsis thaliana.
XX      PN      EPI033405-A2.
XX      PD      06-SEP-2000.
XX      PE      25-FEB-2000; 2000EP-0301439.
XX      PR      25-FEB-1999; 99US-0121825.
XX      PR      05-MAR-1999; 99US-0123160.
XX      PR      09-MAR-1999; 99US-0123548.
XX      PR      23-MAR-1999; 99US-0125788.
XX      PR      25-MAR-1999; 99US-0126264.
XX      PR      29-MAR-1999; 99US-0126785.
XX      PR      01-APR-1999; 99US-0127462.
XX      PR      06-APR-1999; 99US-0128234.
XX      PR      08-APR-1999; 99US-0128714.
XX      PR      16-APR-1999; 99US-0129845.
XX      PR      19-APR-1999; 99US-0130077.
XX      PR      21-APR-1999; 99US-0130449.
XX      PR      23-APR-1999; 99US-0130510.
XX      PR      28-APR-1999; 99US-0130891.
XX      PR      30-APR-1999; 99US-0132048.
XX      PR      04-MAY-1999; 99US-0132407.
XX      PR      05-MAY-1999; 99US-0132484.
XX      PR      06-MAY-1999; 99US-0132486.
XX      PR      07-MAY-1999; 99US-0132487.
XX      PR      11-MAY-1999; 99US-0132863.
XX      PR      14-MAY-1999; 99US-0134218.
XX      PR      14-MAY-1999; 99US-0134219.
XX      PR      14-MAY-1999; 99US-0134221.
XX      PR      18-MAY-1999; 99US-0134370.
XX      PR      18-MAY-1999; 99US-0134768.
XX      PR      19-MAY-1999; 99US-0134941.
XX      PR      20-MAY-1999; 99US-0135124.
XX      PR      21-MAY-1999; 99US-0135353.
XX      PR      24-MAY-1999; 99US-0135629.
XX      PR      25-MAY-1999; 99US-0136021.
XX      PR      27-MAY-1999; 99US-0136392.
XX      PR      28-MAY-1999; 99US-0136782.
XX      PR      01-JUN-1999; 99US-0137222.
XX      PR      03-JUN-1999; 99US-0137528.
XX      PR      04-JUN-1999; 99US-0137502.
XX      PR      07-JUN-1999; 99US-0137724.
XX      PR      08-JUN-1999; 99US-0138094.
XX      PR      10-JUN-1999; 99US-0138540.
XX      PR      10-JUN-1999; 99US-0138847.
XX      PR      14-JUN-1999; 99US-0139119.
XX      PR      16-JUN-1999; 99US-0139452.
XX      PR      16-JUN-1999; 99US-0139453.
XX      PR      17-JUN-1999; 99US-0139492.
XX      PR      18-JUN-1999; 99US-0139454.
XX      PR      18-JUN-1999; 99US-0139455.
XX      PR      18-JUN-1999; 99US-0139456.
XX      PR      18-JUN-1999; 99US-0139457.
XX      PR      18-JUN-1999; 99US-0139458.
XX      PR      18-JUN-1999; 99US-0139459.
XX      PR      18-JUN-1999; 99US-0139460.
XX      PR      18-JUN-1999; 99US-0139461.
XX      PR      18-JUN-1999; 99US-0139462.
XX      PR      18-JUN-1999; 99US-0139463.
XX      PR      18-JUN-1999; 99US-0139750.
XX      PR      18-JUN-1999; 99US-0139763.
XX      PR      21-JUN-1999; 99US-0139817.
XX      PR      22-JUN-1999; 99US-0139899.
XX      PR      23-JUN-1999; 99US-0140353.
XX      PR      23-JUN-1999; 99US-0140354.
XX      PR      24-JUN-1999; 99US-0140695.
XX      PR      28-JUN-1999; 99US-0140823.
XX      PR      29-JUN-1999; 99US-0140991.
XX      PR      30-JUN-1999; 99US-0141287.
XX      PR      01-JUL-1999; 99US-0141842.
XX      PR      01-JUL-1999; 99US-0142154.
XX      PR      02-JUL-1999; 99US-0142055.
XX      PR      06-JUL-1999; 99US-0142390.
XX      PR      08-JUL-1999; 99US-0142803.
XX      PR      09-JUL-1999; 99US-0142920.
XX      PR      12-JUL-1999; 99US-0142977.
XX      PR      13-JUL-1999; 99US-0143542.
XX      PR      14-JUL-1999; 99US-0143624.
XX      PR      15-JUL-1999; 99US-0144005.
XX      PR      16-JUL-1999; 99US-0144086.
XX      PR      16-JUL-1999; 99US-0144086.
XX      PR      19-JUL-1999; 99US-0144325.
XX      PR      19-JUL-1999; 99US-0144331.
XX      PR      19-JUL-1999; 99US-0144332.
XX      PR      19-JUL-1999; 99US-0144333.
XX      PR      19-JUL-1999; 99US-0144334.
XX      PR      19-JUL-1999; 99US-0144335.
XX      PR      20-JUL-1999; 99US-0144352.
XX      PR      20-JUL-1999; 99US-0144632.
XX      PR      20-JUL-1999; 99US-0144684.
XX      PR      21-JUL-1999; 99US-0144814.
XX      PR      21-JUL-1999; 99US-0145086.
XX      PR      21-JUL-1999; 99US-0145088.
XX      PR      22-JUL-1999; 99US-0145085.
XX      PR      22-JUL-1999; 99US-0145087.
XX      PR      22-JUL-1999; 99US-0145089.
XX      PR      22-JUL-1999; 99US-0145192.
XX      PR      23-JUL-1999; 99US-0145145.
XX      PR      23-JUL-1999; 99US-0145218.
XX      PR      23-JUL-1999; 99US-0145224.
XX      PR      26-JUL-1999; 99US-0145276.
XX      PR      27-JUL-1999; 99US-0145913.
XX      PR      27-JUL-1999; 99US-0145918.
XX      PR      27-JUL-1999; 99US-0145919.
XX      PR      28-JUL-1999; 99US-0145951.
XX      PR      02-AUG-1999; 99US-0146386.
XX      PR      02-AUG-1999; 99US-0146388.
XX      PR      02-AUG-1999; 99US-0146389.
XX      PR      03-AUG-1999; 99US-0147038.
XX      PR      04-AUG-1999; 99US-0147204.
XX      PR      04-AUG-1999; 99US-0147302.
XX      PR      05-AUG-1999; 99US-0147192.
XX      PR      05-AUG-1999; 99US-0147260.
XX      PR      06-AUG-1999; 99US-0147303.
XX      PR      06-AUG-1999; 99US-0147416.
XX      PR      09-AUG-1999; 99US-0147493.
XX      PR      09-AUG-1999; 99US-0147935.
XX      PR      10-AUG-1999; 99US-0148171.
XX      PR      11-AUG-1999; 99US-0148319.
XX      PR      12-AUG-1999; 99US-0148341.
XX      PR      13-AUG-1999; 99US-0148565.
XX      PR      13-AUG-1999; 99US-0148684.
XX      PR      16-AUG-1999; 99US-0149368.
XX      PR      17-AUG-1999; 99US-0149175.
XX      PR      18-AUG-1999; 99US-0149426.
XX      PR      18-AUG-1999; 99US-0149722.
XX      PR      20-AUG-1999; 99US-0149723.
```

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 40.8%; Score 608.5; DB 21; Length 302;

Best Local Similarity 50.6%; Pred. No. 1.2e-47;

Matches 131; Conservative 34; Mismatches 65; Indels 29; Gaps 9;

QY 22 DIVGMALP-----GLAPNLSASVSASASEDSAKKIRKPYTIKRSRPSWSEOHDKFLFA 76  
DB 15 dsnmslpsdsdgyfslpatgrts-tvafsedptkikrpytlkksrtenwctgdghdkflfa 73  
QY 77 LQLEFRDMKKIEAFVSGSKTYIQIRSHAQKYYFLKVOKNKGTREHYVPPRPKRKASHPYPOKA 136  
DB 74 lhlfdfrwkkleefvsgsktvtgrlshaqkyflfkvqksanehlpprpkrkashpypika 133  
QY 137 SKNPPVSOQVSTAFPTIRAT--QLDSCGYPPRAESSILTKSGSCPTVSSWVHHHTIRSIDA 194  
DB 134 pkhv-----aytslpsstlpllepgylyssdsksimngncaavcastssw-nheshtlpk 187

QY 195 SFEKDDGGPFCIET-----GNNCSSGSTP-----SSPTMPKCEIPEKVKPDESOYK 244  
DB 188 pvltee-----pvsataplpmnrcrgedteravtkpneesc-ekhrmpntaeyys 242  
QY 245 FIGSVFDPSTTDHLKRLKE 263  
DB 243 flgsvfdpntsglqlqkq 261  
RESULT 6  
AAC45426  
ID AAC45426 standard; Protein; 303 AA.  
XX  
AC AAC45426;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment seq ID NO: 57030.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140635.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142350.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148668.  
PR 16-AUG-1999; 99US-0149364.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 40.8%; Score 608; DB 21; Length 303;  
Pred. No. 13e-47;

Matches 131; Conservative 34; Mismatches 66; Indels 28; Gaps 9;

QY 22 DVGMLP-----GLAPNLSASVSASAPDSAKRKPYTTIKSRSEMSOEHDKFLRA 76  
DB 15 dssmstpsdgdgfgslpatgrts-tvtfseddptklrkpytlkksrenwcdqendkfllea 73





PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145324.

PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-01589293.  
 PR 13-OCT-1999; 99US-0158294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.

Query Match	40.6%	Score 606	DB 21	Length 285
Best Local Similarity	51.0%	Pred. No. 1.9e-47		
Matches 130	Conservative 34	Mismatches 63	Indels 28	Gaps 9
OY	26 MALP-----GLAPNLSSASVASSEDASAKTRRPPYTIIRKSRSEMSSEQHDKFLLEALOLF	80		
DB	1 malpsadgfgslpapatgrts-tvsfseadpttklrkpytlktsstenvtdgndkfllealnlf	59		
OY	81 DRDMKKIEAFVSGSKTVYIQIRSHAKQYFLKLVQKNGTREHVPPRPKRKASHPPOKASKNV	140		
DB	60 drdwkklleaevgsklvvqgrlshagkyflkvqsgsnehlpprrprkkaahpylikapkv	119		
OY	141 PVSQGVSTAFPPAAT--QLDSCGYPRAESSILITKSGSCSPVSSVWHTTIRSIDASPYE	198		
DB	120 ----aYstslpsasatlpllepylyvsdskslmgngavacatssav-nhesnlpkpyie	173		
OY	199 KDDGPGPIET-----GNCCSSGTF-----SSPTWPPCSFIEPKVKRDPFSQVYKFTGS	248		
DB	174 vee----pyvasataplpmncrcrqedcervravtkpmeesc-ekphrvmptlaevysflgs	229		
OY	249 VEDPSTDTHLKLKLE	263		
DB	230 vfdpntsgnlgrlqk	244		
RESULT 9				
AAG36899				
ID	AAG36899 standard; Protein: 284 AA.			
XX	AC AAG36899;			
XX	18-OCT-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 45284.			
KM	Protein identification: signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
XX	terminalisation sequence.			
OS	Arabidopsis thaliana.			
XX	EP1033405-A2.			
PN	06-SEP-2000.			
PD				
PF	25-FEB-2000; 2000EP-0301439.			
XX				
XX	25-FEB-1999; 99US-0121825.			
PR	05-MAR-1999; 99US-0123180.			
PR	09-MAR-1999; 99US-0123548.			
PR	23-MAR-1999; 99US-0125788.			
PR	25-MAR-1999; 99US-0126264.			
PR	29-MAR-1999; 99US-0126785.			
PR	01-APR-1999; 99US-0127462.			
PR	06-APR-1999; 99US-0128234.			
PR	08-APR-1999; 99US-0128714.			
PR	16-APR-1999; 99US-0128845.			
PR	19-APR-1999; 99US-0130077.			
PR	21-APR-1999; 99US-0130449.			
PR	23-APR-1999; 99US-0130510.			
PR	23-APR-1999; 99US-0130891.			
PR	28-APR-1999; 99US-0131449.			

PR	30-APR-1999	9905-01320248
PR	30-APR-1999	9905-01324247
PR	04-MAY-1999	9905-01324884
PR	05-MAY-1999	9905-01324885
PR	06-MAY-1999	9905-01324886
PR	06-MAY-1999	9905-01324887
PR	07-MAY-1999	9905-01324888
PR	11-MAY-1999	9905-01342456
PR	14-MAY-1999	9905-01342418
PR	14-MAY-1999	9905-01342419
PR	14-MAY-1999	9905-01342421
PR	14-MAY-1999	9905-01342420
PR	18-MAY-1999	9905-01347468
PR	19-MAY-1999	9905-01334494
PR	20-MAY-1999	9905-01351244
PR	21-MAY-1999	9905-01355533
PR	24-MAY-1999	9905-01356529
PR	25-MAY-1999	9905-01360621
PR	27-MAY-1999	9905-01367692
PR	01-JUN-1999	9905-01372222
PR	03-JUN-1999	9905-01375728
PR	04-JUN-1999	9905-01377242
PR	07-JUN-1999	9905-01377242
PR	08-JUN-1999	9905-01380694
PR	10-JUN-1999	9905-01385840
PR	10-JUN-1999	9905-01385847
PR	14-JUN-1999	9905-01391199
PR	16-JUN-1999	9905-01394552
PR	16-JUN-1999	9905-01394553
PR	17-JUN-1999	9905-01394542
PR	18-JUN-1999	9905-01394544
PR	18-JUN-1999	9905-01394610
PR	18-JUN-1999	9905-01394611
PR	18-JUN-1999	9905-01394655
PR	18-JUN-1999	9905-01394656
PR	18-JUN-1999	9905-01394657
PR	18-JUN-1999	9905-01394658
PR	18-JUN-1999	9905-01394559
PR	18-JUN-1999	9905-01394660
PR	18-JUN-1999	9905-01394661
PR	18-JUN-1999	9905-01394662
PR	18-JUN-1999	9905-01394663
PR	18-JUN-1999	9905-01397503
PR	18-JUN-1999	9905-01397503
PR	18-JUN-1999	9905-01397503
PR	21-JUN-1999	9905-01396173
PR	21-JUN-1999	9905-01396173
PR	23-JUN-1999	9905-01405991
PR	30-JUN-1999	9905-01412877
PR	01-JUL-1999	9905-01418462
PR	01-JUL-1999	9905-01421544
PR	02-JUL-1999	9905-01420555
PR	06-JUL-1999	9905-01423990
PR	08-JUL-1999	9905-01428083
PR	09-JUL-1999	9905-01428270
PR	12-JUL-1999	9905-01429277
PR	13-JUL-1999	9905-01435342
PR	14-JUL-1999	9905-01435624
PR	15-JUL-1999	9905-01440005
PR	16-JUL-1999	9905-01440865
PR	16-JUL-1999	9905-01440866
PR	19-JUL-1999	9905-01443251
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443335
PR	20-JUL-1999	9905-01443552
PR	20-JUL-1999	9905-01445352
PR	20-JUL-1999	9905-01445884
PR	21-JUL-1999	9905-01446144



PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 28-JUL-1999; 99US-0145319.  
PR 02-AUG-1999; 99US-0145351.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-015138.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161356.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 40.1%; Score 598.5; DB 21; Length 284;  
Best Local Similarity 55.6%; Pred. No. 9,1e-47;  
Matches 133; Conservative 26; Mismatches 53; Indels 27; Gaps 6;

OY 42 ASASEDSAKRKIRKPYITKRSRESMSBOEHDKFLFALQLFDRBWKIEAFVGSKTVOIRS 101  
Db 3 ssseeliskkkirkpylitksreswlependkilealqlidrtwkkilealigsaktvqirs 62  
OY 102 HAQKPELKVKQKNGTREVHPRPKRKASHPYOKASKNPVSSQOASTAPPTATOLDSCY 161  
Db 63 haqkytlkvqsgstgehlpprpkrkaahpyqkahkvql--qypsgtktsesepndpsf 120  
OY 162 YPPAESSTILTKSGSCPVTSSKWHHTIPSIDASFVEKDDG-----GPPSIEIGNCS 214  
Db 121 mtfresssmllms---ptlaaapwtlnaqtisfcpjpkaycstlmetagaganncs 176  
OY 215 SGCTESSPPTWPCSEIPE-----KVKPDSQVYKKGVSVPDPSTTHLTKIKE 263  
Db 177 s--ssentp---fprsnrdardhgnvghslrvlpdlaqvygylgsvrfdpyashnlqklk 231

RESULT 10  
AAE02543  
ID AAE02543 standard; Protein; 302 AA.  
XX  
AC AAE02543;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE A. thaliana transcription factor G254.  
XX  
KW Plant transcription factor; phenotype; sugar sensing characteristic;  
KW transgenic plant; plant yield; growth; germination; photosynthesis;  
KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;  
KW storage organ; metabolism.  
XX  
OS Arabidopsis thaliana.  
OS  
FH Key Location/Qualifiers  
FT Domain 62..106  
FT /note="Conserved domain"  
PN WO200135725-A1.  
PD  
XX  
XX 25-MAY-2001.  
XX  
PF 14-NOV-2000; 2000MO-US31414.  
XX  
XX 17-NOV-1999; 99US-0166228.  
PR 17-APR-2000; 2000US-0197899.  
PR 22-AUG-2000; 2000US-0227439.  
XX

PA (MEND-) MENDEL BIOTECHNOLOGY INC.  
 PA (JIAN/) JIANG C.  
 PA (HEAR/) HEARD J.  
 PA (PINE/) PINEDA O.  
 PA (PILG/) PILGRIM M.  
 PA (ADAM/) ADAM L.  
 PA (RIEC/) RIECHMANN J. L.  
 PA (YUGG/) YU G.  
 PA (SAMA/) SAMAHA R.  
 XX  
 PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann J.L.  
 PI Yu G, Samaha R;  
 XX  
 DR WPI: 2001-335977/35.  
 DR N-PSDB: AAD06644.  
 XX  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 PT for altering the sugar sensing characteristics of plants and increasing  
 PT yield, e.g. corn, potato and cotton plants -  
 XX  
 PS Claim 4: Page 69-70; 151pp; English.

CC The patent relates to polynucleotides encoding 35 plant transcription  
 CC factors which may be used to modify phenotype associated with a plant's  
 CC sugar sensing characteristics and increasing yield when their expression  
 CC level is altered. Sugars are central regulatory molecules that control  
 CC aspects of physiology, metabolism and development. Therefore the cDNAs  
 CC and proteins of the invention are useful for modifying the growth and  
 CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
 CC respiration, starch and sucrose synthesis and degradation, pathogen  
 CC response, wounding response, cell cycle regulation, pigmentation,  
 CC flowering and senescence of plants and for modifying sink-source  
 CC relationships in seeds, tubers, roots, and other storage organs leading  
 CC to an increase in yield. The transcription factor polynucleotides and  
 CC polypeptides may be used to alter the structure and developmental  
 CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
 CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
 CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,  
 CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,  
 CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
 CC tobacco, tomato, watermelon, roseaceous fruits and/or vegetable brassicas.  
 CC The present sequence is an Arabidopsis thaliana transcription factor.  
 CC  
 XX

SO Sequence 302 AA:

Query Match 40.1%; Score 598.5; DB 22; Length 302;  
 Best Local Similarity 50.2%; Pred. No. 1e-46;  
 Matches 130; Conservative 34; Mismatches 66; Indels 29; Gaps 9;

QY 22 DLVGMALP-----GLAPRLSSASVSASASSEDASAKIRKPTTITRSRSWSQEHDKFLA 76  
 Db 15 dsamslpspsdsgfsglpatgrts-tvsfsepbtkirkpytlkxsrnwtqdehdkfla 73  
 QY 77 LQLPDRMKRTKTEAFVSGKTVQIRSHAKYFLKYOKNGTREVHPPPKRKASHPRYOKA 136  
 Db 74 lhlfdtdkklkkaetvgsaktvqtrshqkyllkvqksqanehrlplpkkkashpryika 133  
 QY 137 SKNVSVSOQVSTAFPTAAT--QLDSGYVPRAESSILTKSGSSCPVSSWVHHTPSIDA 194  
 Db 134 pkuv-----aytslpsasclpllepylylsdskslmngdaqvcassav-nheslnlpk 187  
 QY 195 SFVKKDGGCPGCIET-----GNCGSSGSTE-----SSPTMPPCSEIPEKYKPFQSYK 244  
 Db 188 pvldee-----pvasataplpmntrrgedtervratvktkpmneec-ekphrvmpnfaevs 242  
 QY 245 FICGVFPSTTDHLKLKE 263  
 Db 243 flgsavfdpntsgnlqrllq 261

RESULT 11  
 AAE02559

ID AAE02559 standard; Protein; 293 AA.  
 XX  
 AC AAE02559;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE A. thaliana transcription factor G254 homolog, G673.  
 XX  
 KW Plant transcription factor; phenotype; sugar sensing characteristic;  
 KW transgenic plant; plant yield; growth; germination; photosynthesis;  
 KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
 KW cell cycle regulation; pigmentation; flowering; senescence; physiology;  
 KW storage organ; metabolism.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 37..95  
 FT /note= "Conserved domain"  
 XX  
 PN MO200135725-A1.  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 14-NOV-2000; 2000MO-US31414.  
 XX  
 PR 17-NOV-1999; 99US-0166228.  
 PR 17-APR-2000; 2000US-0197899.  
 PR 22-APR-2000; 2000US-0227439.  
 XX  
 PA (MEND-) MENDEL BIOTECHNOLOGY INC.  
 PA (JIAN/) JIANG C.  
 PA (HEAR/) HEARD J.  
 PA (PINE/) PINEDA O.  
 PA (PILG/) PILGRIM M.  
 PA (ADAM/) ADAM L.  
 PA (RIEC/) RIECHMANN J. L.  
 PA (YUGG/) YU G.  
 PA (SAMA/) SAMAHA R.  
 PA  
 PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann J.L.  
 PI Yu G, Samaha R;  
 XX  
 DR WPI: 2001-335977/35.  
 DR N-PSDB: AAD06660.  
 XX  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 PT for altering the sugar sensing characteristics of plants and increasing  
 PT yield, e.g. corn, potato and cotton plants -  
 XX  
 PS Claim 4: Page 112-113; 151pp; English.

CC The patent relates to polynucleotides encoding 35 plant transcription  
 CC factors which may be used to modify phenotype associated with a plant's  
 CC sugar sensing characteristics and increasing yield when their expression  
 CC level is altered. Sugars are central regulatory molecules that control  
 CC aspects of physiology, metabolism and development. Therefore the cDNAs  
 CC and proteins of the invention are useful for modifying the growth and  
 CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
 CC respiration, starch and sucrose synthesis and degradation, pathogen  
 CC response, wounding response, cell cycle regulation, pigmentation,  
 CC flowering and senescence of plants and for modifying sink-source  
 CC relationships in seeds, tubers, roots, and other storage organs leading  
 CC to an increase in yield. The transcription factor polynucleotides and  
 CC polypeptides may be used to alter the structure and developmental  
 CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
 CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
 CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,  
 CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,  
 CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
 CC tobacco, tomato, watermelon, roseaceous fruits and/or vegetable brassicas.  
 CC The present sequence is a homolog of Arabidopsis thaliana transcription  
 CC factor.



PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146366.  
PR 02-AUG-1999; 99US-0146368.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149950.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 38.1%; Score 568.5; DB 21; Length 293;  
Best Local Similarity 46.6%; Pred. No. 5.5e-44;  
Matches 132; Conservative 17; Mismatches 63; Indels 71; Gaps 6;

QY 5 AETTRDSFETTSGGSSVDLVGMALPGLAPMLSSASVSASASEPSAKKIRKPTTITSRES 64  
DB 14 AETSDAETCT-----latleaageapekvrkayltksres 50  
QY 65 WSEOHDKFLFALQLPDRDKKITEAFVGSKTVIQRSHACKYFLAKYOKNCTREHVPVPPR 124  
DB 51 wtegehbkflleaqlfdrdkkkiedfvgsktvigqishbaqkyflkvyqngtlhpypprp 110  
QY 125 KKKASHPPQKASKNVPVSOQVSTAPPATOLDGYYPRAESSTILTKSGSSCPTVSSW 184  
DB 111 krkaahpyppqkaaknaqmslhvsmstfptqlnl-pytlpwwddtsallnl-----avsgv 164  
QY 185 VHHRTIPSIDASFVEKDGGPGIETGNCS-----GSTESSPTWPPCSIPPKV----- 236  
DB 165 I-----ppdelatlcgaevdvgsndisetsaasgigssrslad 206  
QY 237 -----PDFSQVKKFGSVDPSTTDHLKKLKE 263  
DB 207 skglrlakqapmbglpdaevynflgsvfdpdksgymkklke 249  
RESULT 13  
AAB32797  
ID AAB32797 standard; protein; 142 AA.  
XX  
AC AAB32797;  
XX  
DT 25-JAN-2001 (first entry)  
XX  
DE Eucalyptus grandis transcription factor protein sequence #255.  
XX  
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
KW poplar; sweetgum; teak; mahogany; bzlp; g-box binding factor;  
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MAUS;  
KW homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;  
XX type 2 Cys2His2; CCAT box element; MYB.  
OS  
XX Eucalyptus grandis.  
XX  
PN MO200053724-A2.  
XX  
PD 14-SEP-2000.  
XX  
PE 09-MAR-2000; 2000MO-US06112.  
XX  
PR 11-MAR-1999; 99US-0266513.  
PR 18-AUG-1999; 99US-0149485.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
PI Wood M, McGrath A, Shenk MA, Glenn M;  
XX  
DR WPI; 2000-579369/54.  
XX  
PT New isolated polynucleotide encoding a plant transcription factor for

PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide  
PT -  
XX  
PS Claim 8; Page 308; 747pp; English.

Query Match	32.28;	Score 480.5;	DB 21;	Length 142;
Best Local Similarly	76.48;	Pred. No. 2.5e-36;		
Matches 94;	Conservative 10;	Mismatches 18;	Indels 1;	Gaps 1.

QY	151	PTA	153
		::	
Db	131	pss	133

AC AAB33245;

DE Eucalyptus grandis transcription factor protein sequence #402.

05 Eucalyptus grandis.

PN WO200053724-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-US06112.

PR	11-MAR-1999;	99US-0266513.
PR	18-AUG-1999;	99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

DR WPI; 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for

PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT -  
 XX  
 PS Claim 8, Page 692; 747pp; English.

Query Match	32.28	Score 480.5	DB 21	Length 150
Best Local Similarity	76.48	Pred. No. 2.7e-36		
Matches 94	Conservative 10	Mismatches 18	Indels 1	Gaps 1

QY	151	PTA	153
		1::	
Db	139	pss	141

AA AC AAG20007;

DE Arabidopsis thaliana protein fragment SEQ ID NO: 22027

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

Arabidopsis thaliana.

AA  
PN  
EP1033405-A2.

PD 06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264

PR 01-APR-1999; 99US-0127462..

PR 08-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137562.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.



•  
•  
•  
•

,

,

,



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 01:49:35 ; Search time 1594.5 Seconds  
(without alignments)  
7296.566 Million cell updates/sec

Title: US-09-640-211A-2076  
Perfect score: 862  
Sequence: 1 caaacgcctccgttctctc.....gcttaaaaaaaaaaaaaa 862

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	245.4	28.5	500	AI897784	AI897784 EST267227
2	245.4	28.5	592	AM928286	AM928286 EST307029
3	243.8	28.3	519	AI897681	AI897681 EST267124
4	242	28.1	476	BE057370	BE057370 sm02c04.y
5	241.8	28.1	395	BE805071	BE805071 ss37a07.y
6	241.8	28.1	523	AI930997	AI930997 sb45h07.y
7	238.8	27.7	425	AM423958	AM423958 sb58c05.y
8	238.2	27.6	688	BI272897	BI272897 NF091A12F
9	237.6	27.6	558	BE324639	BE324639 NF024C04P
10	237.6	27.6	669	BC457971	BC457971 NF037A10P
11	236.4	27.4	447	BI273011	BI273011 NF097F02F
12	235.8	27.4	681	AI486576	AI486576 EST244897
13	234.8	27.2	409	BE804790	BE804790 ss45f11.y
14	204.8	23.8	370	BC642441	BC642441 EST355917
15	203.2	23.6	371	AI487923	AI487923 EST246245
16	200.2	23.2	782	BE658316	BE658316 GM700005B
17	198.2	23.0	470	AM156542	AM156542 se28d11.y

18	193.4	22.4	727	10	BG607379	BG607379 WHE2471.H
19	189.8	22.0	626	9	AW255388	AW255388 M407 pep
20	188.6	21.9	718	10	BG369720	BG369720 HYSME1002
21	183.4	21.3	366	10	BM135365	BM135365 WHE0457.B
22	180	20.9	510	10	BF325282	BF325282 su20e03.y
23	173.6	20.1	492	10	BI262794	BI262794 NF094E02E
24	170.8	19.8	320	9	AI489912	AI489912 EST248251
25	167.8	19.5	467	10	BE807621	BE807621 ss28c05.y
26	164.6	19.1	588	10	BF649523	BF649523 NF080G02E
27	164.2	19.0	557	10	BM178036	BM178036 sa168c09.
28	162	18.8	635	9	AW032656	AW032656 EST276215
29	161.4	18.7	412	9	AW756015	AW756015 s112e11.y
30	158.4	18.4	452	10	BF425079	BF425079 su41a07.y
31	158.4	18.4	536	10	BI701275	BI701275 ssg56608
32	158.4	18.4	566	10	BI424418	BI424418 srf34e05
33	157	18.2	408	10	BF647022	BF647022 NF026F03E
34	157	18.2	567	9	AW737355	AW737355 EST338878
35	156.8	18.2	686	10	BG320722	BG320722 Zm04_07a0
36	156	18.1	502	10	BE459188	BE459188 EST414480
37	155.4	18.0	734	9	AI055122	AI055122 com0003B
38	154.8	18.0	631	9	AL507948	AL507948 AL507948
39	152.6	17.7	649	9	BE124286	BE124286 EST394411
40	150.4	17.4	715	10	BE658737	BE658737 GM700007A
41	149.4	17.3	442	10	BM356958	BM356958 12111-F11
42	149.4	17.3	601	9	AM672062	AM672062 LG1.354.G
43	148.2	17.2	454	10	BG509767	BG509767 ssg25f10
44	145.2	16.8	256	10	BE805198	BE805198 ssg3d09.y
45	142.6	16.5	600	10	BE489186	BE489186 WHE1075.G

#### ALIGNMENTS

RESULT 1  
AI897784  
LOCUS EST267227 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
DEFINITION c1ED30123, mRNA sequence.  
ACCESSION AI897784.2 GI:11388186  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
tomato.  
Lycopersicon esculentum  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids I: Solanales: Solanaceae: Solanum: Lycopersicon.  
Alcala,J., Vrebalov,J., White,R., Matera,A.L., Visions,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nieman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.  
Generation of ESTs from tomato carpel tissue  
Unpublished (1999)  
On Jul 27, 1999 this sequence version replaced gi:5603686.  
Contact: CGCI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1..500  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="c1ED30123"  
/clone\_jlb="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: pBluescript SK(-); Site-1: EcoRI; Site-2: XhoI; c1ED - Tomato Carpel EST Library. OligodT-primed and



FEATURES		Location/Qualifiers	
Source		1..519	
		/organism="Lycopersicon esculentum"	
		/cultivar="TA96"	
		/db_xref="taxon:4081"	
		/clone="cLED30G23"	
		/clone_lib="tomato ovary, TAMU"	
		/tissue_type="carpel"	
		/dev_stage="5 days pre-anthesis to 5 days post-anthesis"	
		/lab_host="XLI-Blue MRF"	
		/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; cLED - Tomato Carpel EST library. OligodT-pripped and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."	
BASE COUNT		193 a	73 c 113 g 140 t
ORIGIN			
Query Match	28.3%	Score 243.8;	DB 9; Length 519;
Best Local Similarity	74.9%	Pred. No. 3.9e-44;	
Matches 305;	Conservative 0;	Mismatches 102;	Indels 0; Gaps 0;
Qy	13	tttcctccaagctgaacatggaacagaagccagacagacagtgtgaagtcccaagat	72
Db	27	TTTTTTTAAAAAATAATATATAAAAAAAGATGATAAAATATGCAACCTCTAAAGAT	86
Qy	73	gtcgaagtcgaagaaagggccgtgagcagatgtgaagaagatccatccatcaacata	132
Db	87	GTTGAAGTGAAGAAAGGCGCTTGAGCTATGCAAGAAAGATTATTCCTATTACTACTATT	146
Qy	133	gcgaatcacggcgcgaagcagatggaacatccctagaccaagctgctgtctlaaagctaac	192
Db	147	GCTATTCATGTTGAAGAGTGTTTGAACCTCTAGCTAATATCGTGTGCTCAAGCTACT	206
Qy	193	gggaagagttgtcgctccgtgtgctgtgaactatctgcgaaccgcgtccggaagagcaac	252
Db	207	GGAAAAAGCTTGTGAGCTCGCATGCTTAATATCTTCGACCTGAGTGAAGGCGGCTAAT	266
Qy	253	atcacatcagagagcagcgtctcctgcatctgagacatgcagacagctgggaaagacgtg	312
Db	267	ATTACACTGGAAGAACCACTTTTGATTATGCAACTACATGCTTAAGTGGGGAACCAAGTGG	326
Qy	313	tctaaattgcgaagcatcttcccggaagagacatgagataaagaactctggaag	372
Db	327	TCAAAAATTCGCAACCATTTGGCTCGAAGAACGGAATACGAGATTAATAAAACCTACGTGAGG	386
Qy	373	actagaatccaaagacaatcgaagcagaaggtctctctgtca	419
Db	387	ACTAGATTTCAGAACACATTTAACGACAGAGAACATCAATGAAGGACA	433
RESULT 4			
BE057370		LOCUS	
DEFINITION		BE057370 476 bp mRNA linear EST 03-DEC-2001	
ACCESSION		sm20c04.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:201	
VERSION		Gm-c1015-3838.5' similar to FR:049020 049020 MYB-LIKE DNA-BINDING	
KEYWORDS		DOMAIN PROTEIN. ; mRNA sequence.	
SOURCE		BE057370	
ORGANISM		BE057370.1 GI:8401736	
REFERENCE		EST.	
AUTHORS		soybean. glycine max Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine. 1 (bases 1 to 476) Shoemaker,R., Kelm,P., Vodkin,L., Expelding,J., Coryell,V., Rhanna, A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelshing,B., Allen,M., Bowers, Y., Pearson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk, R., Rittler,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann	

TITLE	JOURNAL	COMMENT
R., Waterston, R. and Wilson, R.	Public Soybean EST Project	Unpublished (1999)
Contact: Shoemaker R/public Soybean EST Project	Public Soybean EST Project	Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA	Tel: 314 286 1800	Fax: 314 286 1810
Email: est@watson.wustl.edu	This clone is available through: ResGen, Invitrogen Corp. 2130	South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: cc@resgen.com	Insert Length: 909	Std Error: 0.00
Seq primer: -40RP from GIDCO	High quality sequence stopper	Location/Qualifiers
1. 476	/organism="Glycine max"	/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-8383"	/clone.lib="Gm-cl015"	/tissue_type="Mature flowers, field grown plants"
/host="XLI10-Gold"	/note="Vector: Bluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cdna library was constructed from mRNA isolated from mature flowers of field grown plants. The cdna library was prepared using the Stratagene pluscript II XR cdna library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cdna fragments followed by XhoI digestion. The cdna fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluscript vector. The ligated cdna fragments were transformed into XLI10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg."	
BASE COUNT	145 a	113 c 116 g 102 t
ORIGIN		
Query Match	28.1%	Score 242; DB 9; Length 476;
Best Local Similarity	74.6%	Pred. No. 9.9e-44;
Matches 337; Conservative	0;	Mismatches 100; Indels 15; Gaps 2;
61 aagtcaccaagatgtctcgaggtgagaaagggccggtgagcagatgagagatctatcttc	120	
16 AGCTCTCATGATCTCGAAGTGTAGAAAGGGGCGCATGTGACATGGAAGAACATAATTACTTG	75	
121 atcaactacatagcgaaatcagcgagaaagcgattggaactccctagccaaagctgctgt	180	
76 ATCACCCTATATTGTCGCAATTCACGCGGGAAGGGGTTTGGAACTCTTTGGCCAAAGCTGCTGGA	135	
181 ctaaaacgtaccgggaaagagattgtctcgctcggctggtcgaactatctgagaccgagcgtc	240	
136 CTTAAACGTACCGGAAAGAGTGTCCGGGCTCGGTGGCTTAACCTACCTCGTCTGATGTT	195	
241 cggagagcgacatcaactctctgagagacagctctcgatcatatgagactgcatgcaagttg	300	
196 ACAAGGCGGAAATTTACACCCGAGGACAGCTTTTGATCATGTGAATTCATCATCAAAAGTGG	255	
301 ggaagacaggtggtctcaaaattgcaaaagcctctcccgagagagacagacaaatagaataag	360	
256 GGAAGACAGGTGTTCCAAAAATTCGCAAGCATCTTACCCGGAAGAGCTGATTAATAGATTAAAG	315	
361 aactcttgagagactagaaatccaaaagcacatcaagcaagcagaagcttctc-----	412	
316 AACCTACTGAGAGACAAGAGTGTCCAGAAAGCAGCTTCAAGCAAGCTTCCAGACGTTCCAGCA	375	
413 -ctgttcagagctccgagatgagttgatt-----caagcaagcaagaagcacaatgtctcagc	465	
376 CAGAGTAGTAATTTGAGTAGTAATTTATTCCTCCCAAGCTTGCACTAGGCCAAGCTGTCCACC	435	

QY	466	atgccagagccgcatggaagcctcacaccac	497
DB	436	ATGGCGCAGCCCATAGAAACCTATTCTCCACC	467
RESULT	5		
BE805071			
LOCUS			
DEFINITION	BE805071	395 bp	mRNA
	5837807.1	Gm-cl061	Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
	Gm-cl061-877	5'	similar to: Q39028 Q39028 ATMY2. ; mRNA sequence.
ACCESSION	BE805071		
KEYWORDS	BE805071.1	GI:10236183	
SOURCE	EST		
ORGANISM	soybean.		
	Glycine max		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
REFERENCE	1 (bases 1 to 395)		
AUTHORS	Shoemaker, R., Kelm, P., Vodka, L., Erpelding, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Knocba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Mateson, R. and Wilson, R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com High quality sequence stop: 367. Location/Qualifiers 1..395 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl061-877" /clone_1bp="Gm-cl061" /tissue_type="mature flowers of field grown plants" /lab_host="DH10B" /note="Vector: pluescript II SK+, site_1: EcoRI, site_2: XhoI; The cDNA library was constructed from mRNA isolated from mature flowers of field grown plants for the cultivar Ralden. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."		
FEATURES			
SOURCE			
BASE COUNT	127 a	86 c	103 g
ORIGIN			79 t
Query Match	28.1%;	Score 241.8;	DB 10; Length 395;
Best Local Similarity	81.8%;	Pred. No. 1.1e-43;	
Matches	279;	Conservative 0;	Mismatches 62; Indels 0; Gaps 0;
QY	61	aagtcacagabctcagagcggaagggccgcgycagatgtaagagatcctc	120
DB	33	ACGCTCATGTATCCGAGAGTGGAGAAAGGGCCATCGAACATGGAAGACTTATCTTG	92

OY		121	atgaacatacagaaatccacgaggagaagcaagtggaaattcccaccgaagctgcgtcgt	180
Db		93	ATCACCTATTATGGCCAAATACCGGGAAGGGCTTGTGGAACTCTTTGGCCAAGGCTGTGGA	152
OY		181	ctaacaactaccggaggaagagtgctcggtgccggttgcgcgaacctacttcgcagcccgacgtc	240
Db		153	CTTAACACTACCGCAAGAAGATTGGCGCGCTCCGGTGCGTTAACAATACTAACCTTCGCCTCATGATT	212
OY		241	cggagagggcacatatcactcactcgaggaagcaagcttcctgatcatctygaactcygacatgccaagt	300
Db		213	AGAAAGGGAATATTATACACCAGAGMAAGCTTTTGATCATAAGAACTTCATGCAAGAGG	272
OY		301	ggaacacagttgtctaaaattgcgaacagcatctcccggaagagactygacaatgagataaag	360
Db		273	GGAACACAGCTGTGCCAAAATTGTCCAGAGATCTACCCGGAAGAGCTGANTAAATGACATTAAAG	332
OY		361	aactcttgaggagactagaaatcccaaacacatcaaagaacg	401
Db		333	AACCTACTGAGAGCAAGAAGATCCAGAACACCTCAAGCAAGC	373
RESULT	6			
LOCUS	A1930997		523 bp	mRNA linear EST 30-NOV-2001
DEFINITION	sbs5h07.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:			
KEYWORDS	Gm-cl015-278 5' similar to TR:Q39028 Q39028 ATMYB2.; mRNA sequence.			
ACCESSION	A1930997			
VERSION	A1930997.1 GI:5666961			
SOURCE	Eukaryote: soybean.			
ORGANISM	Glycine max Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;			
REFERENCE	1 (bases 1 to 523)			
AUTHORS	Shoenmaker,R., Keim,P., Vodkin,L., Erpellding,J., Corvett,J.V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pope,D., Harvey,N., Schurk,R., Rittenberg,E., Korn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Wilson,R.			
TITLE	Public Soybean EST Project			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccure@resgen.com Seq primer: -40RP from gibco High quality sequence stop: 422.  Location/Qualifiers 1..523 /organism="Glycine max" /db_xref="taxon:3847" /clone-"GENOME SYSTEMS CLONE ID: Gm-cl015-278" /cloze.lib-"Gm-cl015" /tissue_type="Mature flowers, field grown plants" /lab_host="XL10-Gold" /note-"Vector: plusscript II XR; site_1: EcoRI; site_2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Strategene pluscript II XR cDNA library construction kit. Complementary DNA was synthesized from mrna using a primer consisting of a poly (dT) sequence with a xhoI restriction site. EcoRI adaptters were ligated to the blunt-ended cDNA frgments followed by			



KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

REFERENCE Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Weller, J.W., and May, G.D.  
AUTHORS Flores, H.R., Imman, J.T., Weller, J.W., and May, G.D.  
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula flower library  
JOURNAL Unpublished (2001)  
COMMENT Contact: May, G.D.  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Insert Length: 688 Std Error: 0.00  
Plate: 091 row: A column: 12  
Seq primer: TCACACAGGAACACCTGTGAC.  
Location/Qualifiers

## FEATURES

source

1..688  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF091A12FL"  
/clone\_1lb="Developing flower"  
/tissue\_type="Developing flowers"  
/dev\_stage="Developmentally pooled. Contains a mixture of very young, developing, fully-opened flowers and flowers in early transition into pods."  
/note="Vector: Lambda Zap; cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the GigaPack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using Exassist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

BASE COUNT 235 a 144 c 125 g 182 t 2 others  
ORIGIN

## Query Match

Best Local Similarity 79.2%; Score 238.2; DB 10; Length 688;

Matches 282; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 50 acgacagtgtaagtcgccgaatcgcgaagtcgagaaagggccgtgagcagtcgagaaag 109  
DB 59 AAGAGTCGAGTTCTTCAAGATCCTGATATAGCAAAAGGACCATGACATGGAAGAAAG 118  
QY 110 atccatccatcaatacatagcgaatcagcggaagcgagtcggaatccctcagca 169  
DB 119 ACTTGATCTTGATCACTATATATTCGCCAATCATGNGAAGCTGTTTGAACTCCTTGCTA 178  
QY 170 aagctgctgctcctaagcgaatcgcgaagtcgctcgcgctgcgctggaatcctgc 229  
DB 179 AATCTGCTGGCTTAAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAAC 238  
QY 230 gaccgcagctcgcgagagcacaatcactactgagagcagctcctgataatgaaactgc 289  
DB 239 GACCGGATGTAGACAGAGCAATATTACACCTGAGCAACACTCTGATCATTTGAACCTTC 298  
QY 290 atgcaagtcgagaaacagcgctgctcctaagtcgaaatgcaagcactcctccgagagcagca 349  
DB 299 ATGCTAACTGGGAAACAGGCTGCTGAAATTCGGAACACTTACCGCGGAAGAACGATTA 358  
QY 350 atgagataaagaaactctgagagactagaatccaaagcacaatcaagcaagcagag 405

DB

359 ATGAGATTAAGACTACTGAGAGCAAGATCCAAAGCATATCAAGCAAGCTGAG 414

## RESULT 9

BE324639

LOCUS

DEFINITION

NP024C04P1.F1023 Phosphate starved leaf Medicago truncatula cDNA

clone NF024C04PL 5', mRNA sequence.

Accession

BE324639.2 GI:11965871

VERSION

BE324639.2

KEYWORDS

EST.

SOURCE

barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;

Medicago.

1 (bases 1 to 558)

Lin, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores,

H.R., Imman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation

Medicago truncatula phosphate-starved leaf library

Unpublished (2000)

On Jul 14, 2000 this sequence version replaced gi:1918416.

Contact: Harrison MJ

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7325

Fax: 580 221 7380

Email: mjharrison@noble.org

Medicago genome initiative accession: MGI:S:24661

Insert Length: 698 Std Error: 0.00

Plate: 024 row: C column: 04

Seq primer: TCACACAGGAACACCTGTGAC.

Location/Qualifiers

1..558

/organism="Medicago truncatula"

/db\_xref="taxon:3880"

/clone="NF024C04PL"

/clone\_1lb="phosphate starved leaf"

/tissue\_type="leaf"

/dev\_stage="trifoliolate"

/note="Vector: Lambda Zap; At the trifoliolate stage, M.

truncatula plants were transplanted to phosphate-free sand

and grown for a further 30 days. During this 30 day

period, the plants were fertilized twice weekly with 1/2

Hoaglands solution containing only 20uM potassium

phosphate. RNA was prepared from above ground tissues."

BASE COUNT 196 a 112 c 110 g 140 t

ORIGIN

Query Match

Best Local Similarity 79.2%; Score 237.6; DB 10; Length 558;

Matches 282; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 50 acgacagtgtaagtcgccgaatcgcgaagtcgagaaagggccgtgagcagtcgagaaag 109  
DB 60 AAGAGTCGAGTTCTTCAAGATCCTGATATAGCAAAAGGACCATGACATGGAAGAAAG 119  
QY 110 atccatccatcaatacatagcgaatcagcggaagcgagtcggaatccctcagca 169  
DB 120 ACTTGATCTTGATCACTATATATTCGCCAATCATGTAAGCTGTTTGAACTCCTTGCTA 179  
QY 170 aagctgctgctcctaagcgaatcgcgaagtcgctcgcgctgcgctggaatcctgc 229  
DB 180 AATCTGCTGGCTTAAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAAC 239  
QY 230 gaccgcagctcgcgagagcacaatcactactgagagcagctcctgataatgaaactgc 289  
DB 240 GACCGGATGTAGACAGAGCAATATTACACCTGAGCAACACTCTGATCATTTGAACCTTC 299

QY 290 atgccagctgggggaacagctgtgtctcaaaatttcgaagcatctcccggaagactgaca 349  
|||||  
Db 300 ATGCTAAGTGGGGAACAGCTGTCTCAAAATTTCGAACATCTACCGGGAAGACCGATA 359

QY 350 atgagaataagaactcttgaggactagatccaaaagacatcaagcaagacagag 405  
|||||  
Db 360 ATGAGATTAAGAAGACTCTGTGAGAGACAGATCCAAAAGCATATCAAGCAAGCTGAG 415

RESULT 10  
BG457971 669 bp mRNA linear EST 19-MAR-2001  
LOCUS NF037A10PLF1070 Phosphate starved leaf Medicago truncatula cDNA  
DEFINITION clone NF037A10PL 5', mRNA sequence.  
ACCESSION BG457971  
VERSION BG457971.1 GI:13381296  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 669)  
AUTHORS Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores  
,H.R., Imman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.  
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
JOURNAL Medicago truncatula phosphate-starved leaf library  
COMMENT Unpublished (2000)  
CONTACT: Harrison MJ  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7325  
Fax: 580 221 7380  
Email: mjharrison@noble.org  
Insert Length: 669 Std Error: 0.00  
Plate: 037 row: A column: 10  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1..669  
FEATURES  
source  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF037A10PL"  
/clone\_lib="Phosphate starved leaf"  
/tissue\_type="leaf"  
/dev\_stage="trifoliolate"  
/note="Vector: lambda Zap: At the trifoliolate stage, M.  
truncatula plants were transplanted to phosphate-free sand  
and grown for a further 30 days. During this 30 day  
period, the plants were fertilized twice weekly with 1/2  
Hoaglands solution containing only 20uM potassium  
phosphate. RNA was prepared from above ground tissues."

BASE COUNT 225 a 141 c 127 g 176 t  
ORIGIN

Query Match 27.6%; Score 237.6; DB 10; Length 669;  
Best Local Similarity 79.2%; Pred. No. 9.3e-43;  
Matches 282; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 50 agcagcgtgttaagtcaccaagatgtcgagtgagaaaagggccgtgagcgtgaagag 109  
|||||  
Db 60 AAGAGTCACTTCTTCACAAAGATCTGATGTACGAAAGGACCATGACATGGAAGAG 119

QY 110 atctcaccatcaatcatagcgaatcagcggaagcggttgaaactccctgacga 169  
|||||  
Db 120 ACTTGATCTTGATCACTATATTTGCCATATGATGTAAGGTGTTGGAACTCTTGCGCTA 179

QY 170 aagcgtcgtgtctaaacagctaccggaagagttgtcgcgtccggtgtgctgaactatcgc 229  
|||||  
Db 180 AATCTGCTGGCTTTAAACGTACCGAAAGAGTTGCAAGGCTTCGCTGCTAATATCTTC 239

QY 230 gaccgcagctccggagaggaacacatcactactaggaggagcagctctgtatcatggaactgc 289  
|||||  
Db 240 GACCGGAGTGTAGACAGGAGATATTACCTCGAGGAACAACCTGTTGCAATGGAACCTTC 299

QY 290 atgccagctgggggaacagctgtgtctcaaaatttcgaagcatctcccggaagactgaca 349  
|||||  
Db 300 ATGCTAAGTGGGGAACAGCTGTCTCAAAATTTCGAACATCTACCGGGAAGACCGATA 359

QY 350 atgagaataagaactcttgaggactagatccaaaagacatcaagcaagacagag 405  
|||||  
Db 360 ATGAGATTAAGAAGACTCTGTGAGAGACAGATCCAAAAGCATATCAAGCAAGCTGAG 415

RESULT 11  
B1273011 447 bp mRNA linear EST 18-JUN-2001  
LOCUS NF097F02FL1026 Developing flower Medicago truncatula cDNA clone  
DEFINITION NF097F02FL 5', mRNA sequence.  
ACCESSION B1273011  
VERSION B1273011.1 GI:14882835  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 447)  
AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,  
Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.  
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
JOURNAL Medicago truncatula flower library  
COMMENT Unpublished (2001)  
CONTACT: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Insert Length: 447 Std Error: 0.00  
Plate: 097 row: F column: 02  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1..447  
FEATURES  
source  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF097F02FL"  
/clone\_lib="Developing flower"  
/tissue\_type="Developing flowers"  
/dev\_stage="Developmentally pooled. Contains a mixture of  
very young, developing, fully-opened flowers and flowers  
in early transition into pods."  
/note="Vector: lambda zap: cDNA was prepared from polyA+  
enriched, pooled samples of equivalent amounts of total  
RNA from very young, developing, fully-opened flowers and  
flowers transitioning into pods. The cDNA was  
directionally ligated into the Uni-Zap XR vector  
(Stratagene) and packaged using the giga-pack III Gold  
packaging extracts. Phagemids containing cDNA inserts were  
in vivo excised from the recombinant Uni-Zap XR vector  
using ExAssist helper phage and the E. coli strain  
XL1-Blue MRF' (Stratagene). Excised plasmids were plated  
using SOLR cells."

BASE COUNT 158 a 85 c 119 t  
ORIGIN

Query Match 27.4%; Score 236.4; DB 10; Length 447;  
Best Local Similarity 74.6%; Pred. No. 1.7e-42;  
Matches 297; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY	7	tcctcgcttctctccaaagctgaaacatggaagaaagccagacgagcaagtgtaagtc	66
Db	30	TCCTCAATCTCTCTCTCATATACATATATATAAAATATGATATAAAAACCATGCAATCATCT	89
QY	67	caagaatgctgaagtgagaagaagggcgctggagatggaagagatcctcatcaac	126
Db	90	CAAGATTCCTGAAGTGAAGAAAAGGGCCATGCGACCTGTGAACAAACACTGATTTTAAATCAT	149
QY	127	tacatagcaatcacacggcggaagcgcaagtgtgaactccctagccaagaagtctgctgataaa	186
Db	150	TAAATTCGCAAAATCATGATGTAAGGTGTTGCAATTCCTTACCACCAAGCTCTGCTTAAA	209
QY	187	cgtaaccgggaagatgtgtcgagctccggtgagctgaacatcatgtgaccggaagctccgag	246
Db	210	CGTACAGCAAAAAGTTGCGAGCTTCGATGATGTTAAATCTTCCTGTCACATGTTAGAAAG	269
QY	247	ggcaaacatcacactcgaaggaagcagctccgtgatacgaatgcaatgccaagtggggaac	306
Db	270	GGCAATATTACACCTGAGCAACACTTTTGTATCTGATCTGATCAAGCTGGGGAAT	329
QY	307	aggtgtgtcttaaaatgtcaagaatcttcctcggaagagcctcacatgataaagaactc	366
Db	330	AGGGGCTCCAAATTTCCAAAGCATCTTCCAGGAAGAACTGACATGATTAAGAAATTTT	389
QY	367	tggagactagaatccaaaagcacatcaagcaagcaga	404
Db	390	TGCAGACTGATGATACAAAGCACATTAAGCAAGTTGA	427
RESULT 12			
LOCUS	AI486576	681 bp	MRNA
DEFINITION	EST244897 tomato ovary, TAMU Lycopersicon esculentum cDNA clone		EST 18-MAY-2001
ACCESSION	U000000000		
VERSION	AI486576		
KEYWORDS	AI486576.1 GI:4381947		
SOURCE	EST.		
ORGANISM	tomato.		
	Lycopersicon esculentum		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Malvales; eubeberids I; Solanales; Solanaceae; Solanum;		
	Lycopersicon.		
REFERENCE	1 (bases 1 to 681)		
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,L.E.,		
	Liang,F., Updon,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,		
	C.L., Mierman,M., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,		
	S.D. and Giovannoni,J.		
TITLE	Generation of ESTs from tomato carpel tissue		
JOURNAL	unpublished (1999)		
COMMENT	Contact: CUCI		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Email: <a href="http://www.genome.clemson.edu/orders/index.html">http://www.genome.clemson.edu/orders/index.html</a> .		
	Location/Qualifiers		
FEATURES			
source	1..681		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="CLE06E18"		
	/clone_id="tomato ovary, TAMU"		
	/tissue_type="carpel"		
	/dev_stage="5 days pre-anthesis to 5 days post-anthesis"		
	/lab_host="XL1-Blue MRF"		
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:		
	XhoI; cLED - Tomato Carpel EST library. OligoRT-primed and		
	directionally cloned in vector lambda Zap II with 5'		
	and 3' ends located at the EcoRI and XhoI sites,		
	respectively."		
BASE COUNT	248 a 101 c 145 g 187 t		
ORIGIN			

Query Match	27.4%;	Score 235.8;	DB 9;	Length 681;
Best Local Similarity	67.1%;	Pred. No. 2.3e-42;		
Matches 371; Conservative	0;	Mismatches 167;	Indels 15;	Gaps 2

Oy	81	gagaaagaaggccggtaggcacgatctggaagaaggatctcaccaccacaactaacagcaatca	140
Db	8	GAGGAAAGGGCCCTTGGAAGTATGAGAAGAAAGATTTAATTCATTAACTCAATTTGCCAATCA	67
Oy	141	cggcgaaagcagtttggaaacctccctagccaagtcgtctgtcttaaaacytacccggaagag	200
Db	68	TGGTGAAGGTGTTTGGAACTCTCTTAGTAAATCATCTCGTGGCTCACAACGACTGCAAAAAAG	127
Oy	201	tgttcgagcccggtagctgagaaactatccgcagcccgacgcccgagaaggacaatactac	260
Db	128	TTGTAGACTCCGATGGCTAAATTTATCTTCGACCTGATGTCAGAGGGGTAATATTACACC	187
Oy	261	tgaagagcagctctctgatatcatgaactcgtcatctgcagtgagggaagaacagtgtctaagt	320
Db	188	TGGAAGAACACTTTTGGATTATGAAACTACATGCTAAAGTGGGCAAAACAATGGTCAAAAAT	247
Oy	321	tgaagaagcattctcccggaagaagactgcaaatlgaaataaagaactcttgcggaagactagaat	380
Db	248	TGCGAAGCATTTTGCTCGGAAGACGGATAACGAGATAAAAAACTACTGAGAGCATAGAT	307
Oy	381	ccaaaagcacatcaagaagaagagagctctctctgcgcagagctccgagatgagagctac-	439
Db	308	TCGAAGGCACATTAAAGCAAGAGAACATGAAATGGACAAAGAAATTTCAGACACAAAATAT	367
Oy	440	-----aaagaagcacacaagaagcacatgtccagaatgccaagacgagacgaatggaagccta	488
Db	368	TGATCATCAACAAAGGAAGTAGTAGCAATATATCATAGTTGGTCAACGACGATTAACATAGA	427
Oy	489	----cgactacggccgtagtatctccaaggagaacaacaacatgagagcccttcgcggtgaattc	545
Db	428	GACTTACTCTCCCACTTCTACATATGGAATTTTGGACACATAATTTTCAAGCCCAAAATT	487
Oy	546	gtcgtctgagatcaaatgaagccctacttgagacagtcagagacatcttgtctatcgagttaact	605
Db	488	TCTCAATGAACAAATAGCACACATCTGGAGCATGTGAGAGATATCTGTCATGCATGCATTTGCT	547
Oy	606	caatggagagattga 618	
Db	548	TAAATGAGATTAA 560	
RESULT 13			
BE804790		409 bp mRNA linear EST 06-DEC-2001	
LOCUS		ss45f11.y1 Gm-cl061 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:	
DEFINITION		Gm-cl061-1702 5' similar to FR:Q39028 Q39028 ATMYB2.; mRNA sequence.	
ACCESSION	BE804790		
VERSION	BE804790.1	GI:10235902	
KEYWORDS	EST,		
SOURCE	soybean,		
ORGANISM	Glycine max		
REFERENCE	1 (bases 1 to 409)		
AUTHORS	Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,J., Coryell,V., Khana,A., Bolla,B., Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelsting,B., Allen,M., Bowers,Y., Pearson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ratter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Willson,R.		
TITLE	Public soybean EST project		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Shoemaker R/Public soybean EST project Public soybean EST Project Washington University School of Medicine		



4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available through: ResGen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntville, AL 35801 For further information  
 call: (800)-533-4363 or contact via email: coufresgen.com  
 Insert Length: 1068 Std Error: 0.00.

## FEATURES

source

1..409  
 /organism="Glycyne max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl061-1702"  
 /clone.lib="Gm-cl061"  
 /tissue\_type="mature flowers of field grown plants"  
 /lab\_host="DH10B"  
 /note="Vector: Bluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI; The cDNA library was constructed from mRNA isolated  
 from mature flowers of field grown plants for the cultivar  
 Raiden. Complementary DNA was synthesized from mRNA using  
 a primer consisting of a poly(dT) sequence with a XhoI  
 restriction site. EcoRI adapters were ligated to the  
 blunt-ended cDNA fragments followed by XhoI digestion. The  
 cDNA fragments were directionally cloned into the  
 EcoRI-XhoI restriction site of the Bluescript vector. The  
 ligated cDNA fragments were transformed into DH10B host  
 cells (GibcoBRL). This library was constructed in the  
 laboratory of Dr. Randy Shoemaker."

BASE COUNT  
 ORIGIN

136 a 87 c 97 g 89 t

Query Match 27.2%; Score 234.8; DB 10; Length 409;

Best Local Similarity 79.4%; Pred. No. 3.9e-42;

Matches 278; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 61 aatcccaagatcgcaggtgagaaagcgctgagcagatggaagagatcctcctc 120  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 58 ACCTCTCAATCTCTGACGAGCAAAAGACCTTGGACGATGGAAGACCTTGCTTG 117  
 QY 121 atcaactacatagcgaactcagcgagagcagcttggaactccctagccaaagctgtcgt 180  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 118 ATCAACTATATGCAATATCATGAGGAGGTGTTGGAAATCTTTGGCCAAAGCTGCTGT 177  
 QY 181 ctaaaagctccgggaagatctgcgtccggtgctgtaactatctggaaccgagctc 240  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 178 CTCAAAAGTACCGGAAAGATTGCCGCTAAGTGGCTAAACTACCTCGCTGTGATGTT 237  
 QY 241 cggagagcagacatcactactgagagcagctcctgatactggaactcagcaagtgg 300  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 238 AGAAGAGGGAATATTACACCCGAGGAAACAATTGATCATGAGCTTCAACGCAAACTGG 297  
 QY 301 ggaacacagtgctcctaaattgcaaacatctcccggaagagctgcaatgataaag 360  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 298 GGAAGACAGTGGTCACAAATTTGCCAAGCATCTACCTGTGAGACAGTAAATGAGATCAAG 357  
 QY 361 aactctggagagctagaatccaagaacacatcaagaagcagaagcttt 410  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 358 ATCTATTGGAGACAGGATCCAGAACCATCAAGCAAGCTGAGAACTT 407

RESULT 14

BG642441

LOCUS

EST355917 tomato flower buds, anthesis, Cornell University

LYCopersicon esculentum cDNA clone cTOD1K13 5' end, mRNA sequence.

ACCESSION

BG642441

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

## REFERENCE

AUTHORS

1 (bases 1 to 370)

Van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang

,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman

,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, anthesis

Unpublished (1999)

CONTACT: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>.

## FEATURES

source

1..370  
 /organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="cTOD1K13"  
 /clone.lib="tomato flower buds, anthesis, Cornell  
 University"  
 /tissue\_type="flower"  
 /dev\_stage="anthesis"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; supplier: Tanksley; Flower buds and flowers were  
 taken from greenhouse plants (4-8 wks old, TA496). They  
 were immediately frozen in liquid nitrogen and then  
 size-separated while remaining frozen."

BASE COUNT  
 ORIGIN

128 a 53 c 78 g 11 t

Query Match 23.8%; Score 204.8; DB 10; Length 370;

Best Local Similarity 74.7%; Pred. No. 1.8e-35;

Matches 257; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 13 ttctctccaaagtgcagatgacgaagccagagacagctgtaagtcaccaagt 72  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 27 TTTTTTTAAAAAATAAATTTATAAATAAAGATGATAAATATGCAACTCTCAAGAT 86  
 QY 73 gtctgagtgagaaaagggcgctgagcagatggaagagatcctacatcactacata 132  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 87 GTTGAAGTGAGGAAGAGCGCTTGAGACTATGGAAGAAGATTATATCTATTAATCAAT 146  
 QY 133 ggcactacagcggaagcgatggaactccctagccaagctgctgcttaaacgtacc 192  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 147 GCTTAATCATGCTGAACAGCTGTTGGAAGCTCTAGCTAAATCTGCTGCTCAAAAGTACT 206  
 QY 193 ggggaagattgttcggtcgcgtgagctgaactatctggaacccggaagctcgggaaggaac 252  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 207 GGAAGAAGTTGTAGACTCCGATGGCTAAATTTATCTTGACCTGATGTCAAGAGGGGTAAAT 266  
 QY 253 atactactcgaaggaagctcctgatactggaactcagatgcaaacgtggaagcaagctg 312  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 267 ATTACACCTGAAGAACAACTTTTGTATTGAGAACTCATGACGTAAGTGGGGAACAAGTGG 326  
 QY 313 tctaaatctgcaaacgactctcccggaaggaactgtacaatgagat 356  
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
 Db 327 TCMAAAATTTGCGAAGCATTTTCTGCTGAGAGACGAGATMAAGAGAT 370

RESULT 15

AI487923

LOCUS

EST246245 tomato ovary, TAMU Lycopersicon esculentum cDNA clone

CLEED19F4, mRNA sequence.

ACCESSION

AI487923

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;





Chlu, C.W.; Chng, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huizart, L.  
 Nature 408, 816-820, 2000  
 A:Accession: 686145  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1284 <SPD>  
 A:Cross-references: GB:AF005172; NID:g8920588; PIDN:AF81310.1; GSPDB:GM00141  
 C:Genetics:  
 A:Map position: 1

Query Match	38.6%	Score 576:	DB 2:	Length 1284:
Best Local Similarity	50.0%	Pred. No. 1,2e-35:		
Matches 129:	Conservative 31:	Mismatches 54:	Indels 44:	Gaps 8:
QY	26	MALPG---LAPLSSASVASAS-----EDSAKIRKPYTTKRSRESDEGHRKLEA	76	
DB	1013	MSLPFGNTLPHTATTIPVSIISNRMTSPFEDTKVVRKRYTTKRSRMTTEGHNKLEA	1072	
QY	77	LQLPGRDMKKIEAFVGSKVIVQIRSHAQKFLFKVQKNGTREHVPVPRPKASHPYOKA	136	
DB	1073	LALPFRDMKKIKAFVGSKVIVQIRSHAQKFLFKVQKNGTKEHLPPRPKRKANHPYOKA	1132	
QY	137	SKNNPVSOQVSTAFPLAATQLDSCGYPRARESSILTKSSQCPYASWNHNTPLSDAF	196	
DB	1133	PK-----FTLSSSNALFQHDYLY-----NTNSHPVISTTKRKGVLHCDVSI	1173	
QY	197	---VEKDDGCPQGIETGNNCSGSTESEPPT-----WPCSELPKVKKDFDSQVYKF	245	
DB	1174	PSSVIAKEERG---VSENCSTSSRDKORTRIIVETINDQESCCK-PHRAVAPFAEVLNMF	1228	
QY	246	IGSVFDPSTTDLHKKLKE	263	
DB	1229	IGSVFDPKTTGHVKKRLE	1246	

```

RESULT      3
T48305
Hypothetical protein F9G14.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48305
R:Beyann, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 1 2000
A:Reference number: Z24491
A:Accession: T48305
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <BEV>
A:Cross-references: EMBL:AL62973
A:Experimental source: cultivar Columbia; BAC clone F9G14
C:Genetics:
A:Map position: 5
A:Introns: 65/2, 84/3, 126/1, 193/1, 235/1, 271/3, 291/3
A:Note: F9G14.150

```

	Query Match	38.0%	Score 567.5	DB 2	Length 307
	Best Local Similarity	4.8%	Pred. No. 8.6e-36		
	Matches 133	Conservative 17	Mismatches 62	Indels 85	Gaps 6
QY	5 AETRYDSFETTSQCGSSVDLGMALPGLAPRLASASVSASASSEDSAKIKRPTITTSR	64			
	14 AETSTATETET-----IATTEGGEAPKKRVKRAATITTSR	50			
DB	14 AETSTATETET-----IATTEGGEAPKKRVKRAATITTSR	50			

Oy 65 MSDEDHKFLFALQLFDPDMKKILEAFVGSRTVQIRSHAQKVFLLKVKNGKNGRREHVPPRP 124  
 Db 51 WTGEENHKFLFALQLFDRDMKKILEDVGSRTVQIRSHAQKVFLLKVKNGKNGRNGTLAHVPPRP 110  
 Oy 125 KRRASHPYPOKASKNVPVSOOYSTAFPTAAITDLSGTYPRAESSSILTKSSSCPTVSSW 104  
 Db 111 KRKAHPYPOKASKNQMSTLHWSMSPTQINNLPGYTPWDDPTFALLNT----AVSGV 164  
 Oy 185 VHNHTPIIDASFVEKDDGPPGIGETGNNCS-----GSTESSPTPTW 226  
 Db 165 I-----PPDELDITLGGAGGMRSHLYSIFICEFVDVSGNDMISESP 206  
 Oy 227 PCSETPPKVK-----PPFSQYKFGISVDPPTSTDLKRLKE 263  
 Db 207 SASGIGSSSRFLTDSKGLRLAKQAPSMHGLPFAEYVNFPGSVDPDCKGMKKRLKE 263

RESULT 4  
 T01715  
 Hypothetical protein A\_IG002N01.20 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T01715  
 R:Scheet, P.; Maggil, L.  
 submitted to the EMBL Data Library, June 1997  
 A:Description: The sequence of A. thaliana IG002N01.  
 A:Reference number: Z14407  
 A:Accession: T01715  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-213 <SCH>  
 A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191140; ATSP:A\_IG002N01.20  
 C:Genetics:  
 A:Gene: ATSP:A\_IG002N01.20  
 A:Introns: 76/2; 95/3; 137/1; 192/1

```

Query Match Similarity 33.7% Score 502.5 DB 2 Length 213;
Best Local Similarity 56.4% Pred No. 4,8e+31;
Matches 106; Conservative 20; Mismatches 45; Indels 17; Gaps 5;

QY 22 DLVGNALP-----GLAPNLSSASVSASASEDSAKKIRKPYTTTKSRSESVGEQDHKFLFA 76
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 DSSNMSLPSSDDFGSIPARGRTS--TVSFEDEDPYTKRKPYTTTKSRHEMTDDEHKKFLFA 73
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 77 LQLFPRDMKKIAFVSGKTVIOIRHAAQYFLFKVOKNGRREHPVPPRPKRKASHPYOKA 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 74 LHLFEDDMKKIAFVSGKTVVQIRHAAQYFLFKVOKSGNHNLPVPPRPKRKASHPYPIA 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 137 SKNVPSQGVSTAFPTAAT--QLDSCGYPRAESSSILTKSGSSCPTVSSVWHHTI---P 190
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 134 PKNV-----AYTSLPSSSTLPLEEPYGLSSDPSKSLMGQAVCASTSSSMNHSTNLPK 188
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 191 SLIDASFVE 198
   | : | | | :
Db 189 VIEGSEFD 196

```

```

RESULT      5
T02684
MKB-related transcription factor (CCA1) [imported] - Arabidopsis thaliana
M:Alternate names: DNA-binding protein CCA1; protein F19D11.11
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02684; G84907
R:Rounsley, S.D.; Lin, X.; Kail, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; et al.
Submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC F19D11 genomic sequence.
A:Reference number: Z14698
A:Accession: T02684
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA

```







DB 195 IHDI---TTVNIPDSPDAAA 212

## RESULT 15

B45344  
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)  
C:Species: suid herpesvirus 1  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
C:Accession: B45344  
R:Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzler, M.  
Virology 179, 365-377, 1990  
A:title: Pseudotables virus immediate-early gene overlaps with an oppositely oriented ORF  
A:Reference number: A45344; MUID:91021039  
A:Accession: B45344  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1733 <VLC>  
A:cross-references: GR:M3651; NID:9334070; PIDN:AAA47471.1; PID:9334072  
C:superfamily: pseudotables virus 1 nuclear antigen

Query Match 7.8%; Score 116.5; DB 1; Length 1733;  
Best Local Similarity 26.8%; Pred. NO. 1.6;  
Matches 38; Conservative 15; Mismatches 56; Indels 33; Gaps 6;

QY 116 REHVPPRPKRKASH--PYOKASKNV--PVSQOVSTAFPTATQLDGYY-----PRA 165  
DB 93 RDHLPTKTRTRDQHRRPTTTTITIKDQHQHDPILLPTKTLQEDPHLLRPRDPPSA 152  
QY 166 ESSSILTKSGSSCPTVSSWVHTIPSIDASFVEKDDGPGIETGNCSGSGTESSPPT- 224  
DB 153 KTHHHODPPGCGPPSTSSHHHODP-----PCGPPSPPPRPSTSSSSHQGPST 203  
QY 225 -----WPCSEIPEKV 235  
DB 204 RPPPPQRPPRPWPPPS--PQKI 223

Search completed: August 12, 2002, 10:11:07  
Job time: 245 sec